Mastering Audio

the art and the science

third edition

Bob Katz

BECOME A MASTER OF AUDIO

This best-selling book has shown thousands of mixing and mastering engineers, musicians, A&R, and producers how to create great sound. It fully integrates the technical and artistic aspects of mastering and contains new and critical information for professionals and students. Don't leave for the studio without it!

WHAT IS MASTERING?

Mastering is the last creative step when producing a record album or single for disc, home music server, iPod™, broadcast or internet delivery. Bob Katz unravels the technical mysteries that challenge audio engineers daily – in an easy-to-grasp, holistic manner. Learn new and powerful artistic techniques for stereo and surround sound. To master audio you must become a master of audio!

• Sequencing, leveling, and processing • PQ coding and master preparation • How to make a master radio- and Internet-ready • Mixing vs. mastering • Advanced compression and equalization techniques for all styles of music • New insights on high sample rates • Jitter, clocking • Dither, wordlengths • Loudness management • Room calibration, acoustics • Equipment interconnection • More!

"*Bob Katz's well thought out book on mastering is a welcome addition to anyone that works in the production of music and who wants to understand all aspects of this so called mysterious stage of music production... There is enough information here for filling in gaps that even seasoned mastering engineers might have.*"
BERNIE GRUNDMAN, BERNIE GRUNDMAN MASTERING
HOLLYWOOD, CA

"*Bob is a master of the technology changeover from analog to digital. His book covers areas that no others have touched.*" GEORGE MASSENBURG, PRODUCER/ENGINEER
MONTREAL, CANADA

"*An excellent reference for anyone interested in CD Mastering. I don't know of another single source with as much detailed information on the mastering process. Even industry veterans are guaranteed to pick up something they hadn't known or were unsure of.*"
TED JENSEN, CHIEF ENGINEER, STERLING SOUND, **NEW YORK CITY**

"*When I first picked up this book, I couldn't put it down until I had read it all! This book should be required reading for all audio professionals - and not just in mastering. Every studio owner and engineer needs to know about this stuff. Even more so with the third edition's chapters on loudness measurement, normalization, and much more!*" MIKE COLLINS, COMPOSER, AUTHOR
PRO TOOLS 11: MUSIC PRODUCTION, RECORDING, EDITING AND MIXING
LONDON

"*The first piece of equipment [you] should buy is Bob Katz's* Mastering Audio: The Art and the Science." ROGER NICHOLS, ROGER NICHOLS MASTERING
MIAMI

"*In addition to updated techniques, this new edition is organized with a "how to" section for beginners followed by a comprehensive reference section to explain the many "whys." This will make a superior textbook for formal classes.*" BOB OLHSSON, MASTERING ENGINEER
AUDIO MASTERY, **NASHVILLE, TN**

"*Bob presents complex topics in an easily accessible format appealing to beginning and experienced readers alike. Many will be eager for the new content, especially loudness measurement and metering, one of the most crucial subjects for audio engineers working today.*"
GARRETT HAINES, MASTERING ENGINEER, TREELADY STUDIOS
PENNSYLVANIA

"*Bob Katz is a true Jedi Knight of Audio*"
A.T. MICHAEL MACDONALD, MASTERING ENGINEER, ALGORHYTHMS
NEW YORK CITY

"*Master This Book!*"
GLENN MEADOWS, MASTERING ENGINEER, **NASHVILLE**

"*When Bob talks, engineers listen; the third edition of* Mastering Audio *will be a must-read.*"
JOHN ATKINSON, AUDIO PRODUCER, ENGINEER
EDITOR, STEREOPHILE MAGAZINE, **NEW YORK CITY**

Focal Press
Taylor & Francis Group
www.focalpress.com

Third Edition

an informa business
ISBN 978-0-240-81896-2
9 780240 818962

Mastering Audio

the art and the science

Bob Katz

third edition

Focal Press
Taylor & Francis Group
www.focalpress.com

First published 2002 by Focal Press

This edition published 2015 by Focal Press, 70 Blanchard Road, Suite 402, Burlington, MA 01803

and by Focal Press, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Focal Press is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data

Katz, Robert A.

Mastering audio : the art and the science / Bob Katz. -- Third edition.

pages cm

1. Mastering (Sound recordings) 2. Sound--Recording and reproducing. 3. Sound--Recording and reproducing--Digital techniques. I. Title.

TK7881.4.K38 2014

621.389'3--dc23

ISBN: 978-0-240-81896-2 (pbk)

Printed in Canada

Dedication

This book is dedicated to *Mary Kent*, my best friend (and wife) since 1984, without whose love and support I would be absolutely nowhere today!

I also dedicate this book to those mastering engineers whose work I especially admire: *Bernie Grundman, Ted Jensen, Bob Ludwig, Glenn Meadows, Bob Olhsson,* and *Doug Sax.* Your fine work has brought great pleasure into my life through the records you have mastered.

Next, I dedicate this third edition to the upcoming generation of audio engineers... You are about to enter an exciting new world, a revolution which has just begun.

— B. K.

"Bob Katz was one of the first engineers to record music with higher definition than the CD's 16 bits, helped reawaken the world of audio engineering to the sonic beauty and honesty of recording with purist microphone techniques, and has been a leading voice for recordings preserving musical value in the "Loudness Wars." When Bob talks, engineers listen; the third Edition of *Mastering Audio* will be a must-read."

John Atkinson, Editor, Stereophile Magazine, Audio Producer, Engineer, Author, www.stereophile.com

"My experiences with Bob's mastering have been very special. I cannot imagine or consider my project done without Bob's special touches. He is really the master of mastering. He has taken every one of my mixes and brought them to a level of polish and beauty that fulfills and exceeds my imagination. He truly takes my music to the ultimate possible level. I am anxious to read the third edition of Bob's new book, *Mastering Audio.* I think it will help me learn about what Bob does to produce those ultimate masters. You should do the same."

Mano Murthy, Producer, Composer, Music Director. Award-winning Film Composer www.manomurthy.com

"Bob Katz has made some of the best sounding recordings the audio world has ever heard. They are cherished throughout the world. Through his vision and efforts he has championed the quest to recreate the live experience, real musicians in a real space."

David Chesky, Composer, Pianist, Producer, CEO HDtracks, Chesky Records. New York, NY, www.hdtracks.com

"Bob possesses a very special gift. He is able to breathe new life into my projects and mixes by turning them into richer, fuller, more polished versions of themselves. To use a visual analogy, he turns my snapshots into beautiful cinematography. And he does so with the utmost care and musicality. I've appreciated, over the years, how generous Bob has been in sharing his knowledge. His insight has made me a better engineer myself, and has opened me up to a realization of just how deeply one can delve into the realm of sound. The Third Edition of *Mastering Audio* is an example of Bob's love for sound and his willingness to share his craft with others. This book will be my essential reference for keeping up with all things audio!"

John Mock, Composer, Producer, Engineer, Arranger, Studio Musician. Nashville, TN, www.johnmock.com
Credits: Musical work with The Dixie Chicks, James Taylor, Nanci Griffith, Maura O'Connell, many more.

"Bob is a pioneer in the mastering community and in audio education. His work is first class and centered around the philosophy of making music sound its best – not just winning some loudness war. His third edition of *Mastering Audio* is an essential read for anyone involved in audio."

Mark Hornsby, Dir. of Music Production & Artist Relations, Sweetwater Studios, www.sweetwaterstudios.com/

"Bob Katz is the Guru of Mastering. By reading his book you'll be introduced to the secrets that will allow you add the best tone, clarity, warmth and transparency to your already good mixes."

Juan de Marcos Gonzalez, Producer, Arranger. Founder of Sierra Maestra, The Afro-Cuban All Stars, Buena Vista Social Club – Credits: Ibrahim Ferrer, Rubén González, Compay Segundo, more...
www.afrocubanallstarsonline.com

"Bob elevated my new Christmas album to a whole new level with sheer perfection. His World-Class mastering blew me away. To learn how he works his magic, I urge you to read his new book."

Natalie Toro, Broadway and TV actress, Singer, www.natalietoro.com
Credits: A Tale of Two Cities, Les Miserables, In The Heights, Law and Order, Person of Interest, many more

"Audio art and science are inextricably linked. Bob deftly interweaves art and the science making both live and breathe in the third edition of *Mastering Audio.* There is much new and essential information for everyone involved in our profession, including critical information on Loudness Normalization technology now in effect in streaming and broadcast which has the potential to restore full dynamic range to pop record production."

Alan Silverman, Recording Engineer, Mastering Engineer, www.arfmastering.com.
Adjunct Professor, Clive Davis Institute of Recorded Music, New York University. Credits: Album-of-the-Year Grammy Nominee. Norah Jones, Leonard Cohen, Judy Collins, Michael Bublé, Keb Mo, The Kinks, Vanessa Williams, more...

"In my days as a newbie recording engineer, I discovered that understanding the mastering process was crucial on my quest to producing world-class sound. The more time spent with an experienced mastering engineer listening to my projects, the more improvement in my subsequent recordings. Time spent reading the Third Edition of *Mastering Audio* is equivalent to sitting with a world-class mastering engineer who is willing to impart with valuable wisdom, experience and knowledge. The book's newly-revised chapter, *Earientation*, gives the reader a head-start in developing the aural vocabulary and the ear to recognize improvements that can be applied to the final product. As I've said to clients, "Anything is possible, with enough time and enough money." With the Third Edition of *Mastering Audio*, you can understand what is possible."

Jim Anderson, Professor, Clive Davis Institute of Recorded Music, Tisch School of the Arts, New York University, www.tisch.nyu.edu. Past-President - Audio Engineering Society, Grammy-winning Recording Engineer

"Bob Katz's attention to detail, wisdom in recording techniques, and high level of musicianship bring greater life, depth, and dimension to everything I produce. Every mastering session with Bob is both a joy and an education!"

Charlie Bertini, Producer. President, AppleJazz Records, Orlando, FL, www.applejazz.com
Credits: John Allred Quartet, Bill Allred's Jazz Orchestra, Ronnie Leigh, more...

"I have been using Bob Katz's *Mastering Audio* as a textbook in my classes since the first edition came out in 2002, because even though the main subject is the preparation of the final master, the book is a thorough course in Audio Technology (science and aesthetics), indispensible for those interested in mastering the field of audio. Bob Katz has the unusual ability to explain in writing the most complex subjects with clarity and precision, and with plenty of wit."

Raul Valery, Program Chair, Sound & Music Technology, Valencia College, Orlando, FL, www.valenciacollege.edu
Credits: Aldemaro Romero, 101 Strings Orchestra, Guy Saint-Clair & The London Symphony Orchestra, Buddy DeFranco, Orquesta Sinfónica Venezuela, Orquesta Sinfónica Nacional (México)...

"Master This Book!"

Glenn Meadows, Mastering Engineer, Mayfield Mastering, Nashville, TN, www.mayfieldmastering.com
Credits: Dan Fogelberg, Jimmy Buffett, Shania Twain, Steely Dan, many others.

"This new 3rd Edition of Bob Katz's seminal writing on *Mastering Audio* is must reading for anyone practically involved in the creation of recorded music, and in particular everyone who is in the formative stages of learning the art and science of audio. Bob has the gift of addressing the reader as an inspiring mentor, exposing and explaining thoroughly all crucial and delicate aspects of work with sound that could enhance the value and presentation of recorded sound. The book has been completely rewritten and augmented with new material addressing the critical areas of the fast-changing technology and applications. Its layout reflects the workflow of the mastering process and focuses on learning both basics and the advanced aspects of theory and practice. Attentive readers will feel more confident and capable when approaching all sorts of technical and practical challenges in the studio, and will likely keep this book at a close range to use as a reference manual. The topics are beautifully illustrated, clearly laid out, and worded with clarity and simplicity. Many useful practical tips underlying Bob's vast experience are presented, with hints, valuable reminders and case-study examples added, to aid the reader's enjoyment in acquiring the knowledge. Bob's emphasis on training of listening skills comes across in many instances and his *Earientation* chapter offers numerous listening examples of exercises to form in a reader a strong foundation for a lifelong activity. I highly recommend this book to anyone who takes working in audio seriously, and especially to educational institutions dedicated to audio as an invaluable didactic tool."

Wieslaw Woszczyk, James McGill Professor, Director, Recording Studios, Schulich School of Music, McGill University, Montreal, Canada. Past-President, Audio Engineering Society

"Understanding the fundamentals of musical sound: pitch, chords, rhythm, etc.; these are the most important things an audio engineer needs to know. Owning and understanding these concepts will open many doors for you that a plugin cannot. Bob's book will help take you to that next level."

Mark Everton Gray, Studio Engineer, The Palms, Las Vegas, NV. www.palms.com/music-venues/recording-studio.
Credits: Imagine Dragons, Joe Bonamassa, Mega Genesis, The Killers, many more...

"Bob Katz's well-thought-out book on mastering is a welcome addition to anyone that works in the production of music and who wants to understand all aspects of this so called mysterious stage of music production... There is enough information here for filling in gaps that even seasoned mastering engineers might have."

Bernie Grundman, Mastering Engineer. Bernie Grundman Mastering, Hollywood, CA.
www.berniegrundmanmastering.com/ Credits: Multiple Grammy-winning, TEC award-winning, Dr. Dre, Joni Mitchell, Gato Barbieri, Jennifer Warnes, Jackson Browne, many more...

"Bob Katz is a true Jedi Knight of Audio"

A.T. Michael MacDonald, Mastering Engineer, AlgoRhythms, New York City, www.algorhythms1.com
Credits: Ben Allison, Fred Hersch, Matt Wilson, Andrew Hill, more...

" An excellent reference for anyone interested in CD Mastering. I don't know of another single source with as much detailed information on the mastering process. Even industry veterans are guaranteed to pick up something they hadn't known or were unsure of."

Ted Jensen, Chief Engineer, Sterling Sound, New York City, www.sterling-sound.com
Credits: Kings of Leon, Green Day, Alice In Chains, Arcade Fire, Pat Metheny, more...

"When I first picked up this book, I couldn't put it down until I had read it all! This book should be required reading for all audio professionals - and not just in mastering. Every studio owner and engineer needs to know about this stuff. Even more so with the third edition's leading edge approach to loudness measurement, structured advice on how to deal with the advent of loudness normalization in all our mixing and mastering practice. And much more, including a beautiful new layout!"

Mike Collins, Producer, Arranger. Author of *Pro Tools for Music Production*, London, www.mikecollinsmusic.com
Credits: Light of the World, Sun Palace, Juliet Roberts, David 'DaPaul' Philips, more...

"The first piece of equipment [you] should buy is Bob Katz's *Mastering Audio: The Art and the Science."*

Roger Nichols, Audio Engineer, Producer, Mastering Engineer
Credits: Steely Dan, John Denver, Frank Zappa, Stevie Wonder, Crosby Stills & Nash, more...

"With the impending demise of CDs and the rapid expansion of lossy-encoded audio and video music delivery, the requirements of mastering have changed a great deal. In addition to updated techniques, this new edition is organized with a "how to" section for beginners followed by a comprehensive reference section to explain the many "whys." This will make a superior textbook for formal classes..."

Bob Olhsson, Mastering Engineer. Audio Mastery, Nashville, TN, www.audiomastery.com
Credits: Marvin Gaye, The Temptations, Quicksilver Messenger Service, Grateful Dead, more...

"This new edition is essential reading for anyone who wants to keep up with the latest developments in audio production. Bob presents complex topics in an easily accessible format that will appeal to beginning and experienced readers alike. Many will be eager for the new content, especially loudness measurement and metering, one of the most crucial subjects for audio engineers working today."

Garrett Haines, Mastering Engineer, Writer. Treelady Studios, Pennsylvania, www.treelady.com. Senior Contributor, Tape Op Magazine, www.tapeop.com. Credits: Jolie Holland, Watermelon Slim & the Workers, Garage a Trois

"Bob is a master of the technology changeover from analog to digital.
His book covers areas that no others have touched."

George Massenburg, Producer, Engineer. Associate Professor of Sound Recording, Schulich School of Music, McGill University, Montreal, Canada. Grammy-winning, TEC award-winning. Credits: Lyle Lovett, Linda Ronstadt, Little Feat, Earth, Wind & Fire, James Taylor, many others...

Edited by
Christopher Morgan

Foreword
Bob Ludwig

Foreword to the First Edition
Roger Nichols

Contributing Engineers Surround Chapter
Dave Glasser
Morten Lindberg
Bob Ludwig
Rich Tozzoli
Jonathan Wyner

Graphic Design, Layout Typography, Photography
Mary Kent

credits and thanks to...

My friend (of 40+ years) editor *Chris Morgan*, who has helped to make this third edition sound clear, big and warm! *Jim Johnston* for some powerful new third edition diagrams explaining clipping, numerous contributions to facts and figures, fact-checking in this third edition and past editions. *Dr. Uli Brueggemann* for fact-checking in the sections on acoustics and room correction. However, I am the only one to blame for any er3ors that may remain.

My production team: *Mary Kent* has outdone herself with the third edition's beautiful new graphic design, layout, cover design, and her photo-art bringing life to each chapter. *Jim Kaiser*, far more versatile than his title as copy editor, brought the real world experience of teaching audio and mastering at the college level, offering several suggestions. *Sara Brown* did a fabulous job helping me organize about a thousand notes. Oops, Bob, you missed a spot...

Sadly, *Roger Nichols* passed away before the third edition was published. In tribute to his life and contributions to quality audio I asked *Bob Ludwig* to write an updated Foreword. *Gail Kent* for inventing the punny title! *Charlie Bertini*, for finding and preserving a perfect print of the Carnegie Hall chart attached inside the front cover of this book. It has found its way to many studio walls.

Bob Ludwig, one of the busiest and nicest guys in audio, read the first edition manuscript, provided valued initial input. *Dave Glasser, Bob Ludwig, Rich Tozzoli*, and *Jonathan Wyner* updated their surround sound wisdom, and are now joined by *Morten Lindberg's* unique approach for this third edition in Chapter 11.

Rudi Ortner, whose master's thesis and thorough research provided factual documentation of the loudness war for the first time. Rudi's analyses have confirmed my previous estimates and extrapolations and revealed much more. Now everyone can be aware of the facts: It's an astounding story, revealed to the audio world in sometimes grim detail in Chapter 17. Let's do everything we can to prevent another war. Never again.

Konrad Strauss, who reviewed the first edition and made helpful suggestions. *Richard Hulse* for refining the parallel compression technique in the analog domain. *Bob Orban* and *Frank Foti*, radio processing gurus, for providing the text of their excellent article, in Appendix II. Thanks to San Francisco's *Tardon Feathered* and *Marvin Humphrey* of Mr. Toads, for producing the *What is Hot* CD, collaborating and organizing the *What is Hot* competition on the mastering webboard, and for locating Bob Orban to answer that essential question.

B.J. Buchalter of Metric Halo Labs, for devising *SpectraFoo*, one of the most powerful audio analysis tools in the universe, and contributing some important facts. Special thanks to *Bruno Putzeys* for helping on jitter and monitor specifications. *Waves* for their excellent graphics and for making some of the first dynamics processors that go "both ways." Several images in the chapters are adapted from screenshots of Waves products. *Christian G. Frandsen, Paul Frindle, Eelco Grimm, Dan Lavry, Bob Stuart* and *Uli Brueggemann* for technical advice.

Many mentors over the years, including the late *Michael Gerzon, Deane Jensen, Julian Dunn* and *Steve Washburn. Thomas Lund*, one of the most dedicated and artistic engineers around. Thomas edited earlier versions of the manuscripts that evolved into a couple of these chapters. He also provided an image with research into dynamic range tolerance and plenty of good technical advice on loudness matters as well as a formal measurement of iTunes' nominal target level. My colleagues in the Music Loudness Alliance: *Florian Camerer, Eelco Grimm, Kevin Gross, Bob Ludwig, and Thomas Lund. Florian* for being the best diplomat in the audio world and inventing the term, "The Loudness Revolution." My ex-assistant *Robin Reumers*, who is always ready to lend a hand or provide needed audio information. *Mike Chafee*, who listens critically, has taught me a lot about acoustics and helped to improve my monitor system. *Eric James*, editor of the first and second editions, whose sonic signature still rings through these pages!

Megan Ball and the team at *Focal Press*, who exemplify as much care and attention to detail in publishing as we do in audio!

Finally, all my Internet comrades. Your evocative adages, printed with your permission, are found periodically throughout this book.

Fact Check
(some technical sections)
Dr. Uli Brueggemann
Paul Frindle
Jim Johnston
Rudi Ortner
Dick Pierce (2nd ed.)
Bruno Putzeys

Copy Editor
Jim Kaiser

Production Assistants
Sara Brown
Daniel Medina
Christian Robinson
Todd Hays
Ricky Landaeta

Some product photos provided by the manufacturers

contents

Part III: Advanced Theory and Practice

Appendix

Foreword – Third Edition

I've been lucky to have a career mastering music: it has brought me tremendous enjoyment and satisfaction. I've worked with tubes, transistors, lacquer masters for vinyl, 1/4", 1/2", then 1"-wide stereo analog tape, cassettes, digital audio, CDs, mini-discs, surround sound, DVD's, Blu-ray Disc, digital downloads, and now Mastered for iTunes 24-bit sourced AAC encodes that sometimes sound closer to the original 24-bit master than the 16-bit lossless CD.

Even if you own a copy of *Mastering Audio: The Art and Science*, it's time to learn new things again! Bob Katz has re-written the book with new chapters, new information on every page, and it has been reorganized and re-fact-checked by audio authorities, because so much has changed in the audio world!

Recently Spotify, iTunes Radio, and other streaming sites have become popular, and they all use algorithms that turn the level of loud CDs down and bring soft CDs up, making the old "loudness wars" meaningless. Engineers and producers need to know about the *once-in-a-lifetime paradigm shift* from *peak*-level normalized to *loudness*-normalized music. *Mastering Audio's* 3rd Edition has three brand new chapters on this exciting news that could change the way music is produced from now on. Why ruin the sound of your music when it will now be played on internet radio that will keep turning your music down the more you turn it up? After they turn it down, all that distortion and puny sound will sound horrible compared to music with wider dynamics.

This new paradigm is just one reason why this Third Edition of *Mastering Audio* is very important for audio engineers. These new chapters explain the developments in loudness normalization and how they impact every audio engineer's practices and our careers.

You need to read this book – it's up-to-date, factual and it will help take you to a new golden age of audio production and mastering. There will be a time, very soon, where engineers and producers will have the freedom to produce music with any kind of processing they wish; a time where mastering engineers can help the artist create their musical vision without the pressure of competitive loudness altering that vision.

Bob Ludwig
Gateway Mastering & DVD
Portland, Maine, May 2014

Foreword – Previous Editions

When a recording artist I produced heard a great song on the radio he would turn to me and say, "I was going to write that song!" After reading this book my reaction was, "I was going to write this book!" Well, I am glad Bob beat me to it because it looks like he did a much better job than I could have.

What places this book head and shoulders above the rest is the attention to useful detail. Instead of some hyperbole, the reader can actually put these methods to good use. The descriptions of how to perform a task are augmented with the reason that you should perform the task. Not just how downward compressors work, but when and why you would want to use them. Science is meaningless without art.

How do I tell if the digital signal is 16-bit or-24 bit? What does noise shaping do? Should I mix at 96 kHz? How do you make something 3 dB louder when it is already lighting up the over lights? Should I mix to analog or digital? How do I set up my speakers for mixing surround? Which weighs more, a pound of gold or a pound of feathers? These are some of the questions that Bob answers in a clear and concise style.

Bob enters each mastering session with his eyes wide open. Each project is unique, and each mastering session will require a unique approach to bring out the very best results. Bob's musical background helps him select the proper tools for the job. Knowing that a string quartet record does not require the same approach as the Back Street Boys record is half the battle.

Every day clients ask for louder and louder CDs when they come to a mastering session. It is very hard to find Hi-Fidelity CDs these days. Now that you can do your own recording to a digital workstation, buy your own multi-band compressors and burn your own CDs, who needs mastering? My answer is that if you record your own projects at home, you need mastering more than the producer who works with the top engineers in the top studios. The key is outside reference. No, I don't mean that your neighbor came over and said, "Hey, that sounds really great!" I mean reference to other projects, and reference to other engineers who have worked on great sounding CDs.

Bob does an excellent job of dispelling the myth that the louder you make your CD, the louder it will be on the radio. Read this part more than once. Once the reality sinks in, then maybe we will have more viable candidates for a Best Engineered CD Grammy, instead of having to choose a CD for the Least Offensive Engineering award.

The professional mastering engineer works on material from all corners of the music business. This is the last stop before the CD hits the radio and the record stores. The smartest thing any mixing engineer can do is leave the final loudness tools to the loudness professional.

Limiters and compressors should be treated just like firearms. There should be guides for the proper use and classes you must take before you can own one. That class is here in this book. After you read this "audio firearms manual" you will have a much better understanding of the mastering process. You will know when and how to use these tools yourself and when to leave it to the professional. Treat every compressor/limiter as a loaded weapon, and don't point it at anyone unless you intend to use it. It's the **LAW!**

I get e-mail quite often from independent artists who are recording their music at home and want to know what gear to buy to help them mix before they send it to me for mastering. I tell them that the first piece of equipment they should buy is Bob Katz's *Mastering Audio, The Art and the Science.*

Roger Nichols

Miami, August 2002

Roger Nichols passed away April 9, 2011. He never got the chance to enjoy the Loudness Revolution, but I know he's still listening!

Introduction

What Is Mastering?

Mastering is the last creative step in the audio production process, the bridge between mixing and replication (or distribution). It is the last opportunity to enhance sound or repair problems within an acoustically-designed room — under an audio microscope. Mastering engineers lend an objective, experienced ear; we are familiar with what can go wrong technically and esthetically. Sometimes the only process we do is — nothing! The simple act of approval means the mix is esthetically ready for pressing, it only needs to be technically converted to the release medium and proofed. Other times we may help the producer work on the problem song they just couldn't get right in the mix, or add the final touch that makes a record sound finished and playable on a wide variety of systems.

Regardless of the form in which product is sold, our job as mastering engineers remains: we help music to be presented in the best possible way. This requires old-fashioned craftsmanship and attention to detail, values which never go out of style. The artistic and technical information provided in this book will always be precious to students of the art of audio.

On Language

Sex is good! And being sexy can be fun! I feel that language should be sexy, too, and our centuries-old male-centric language must be rather wearying to the women in our society. It's time to put some vitality back into our syntax. Thus, you will find that in one chapter of this book, the Mastering Engineer may be a woman, and in another, a man! Vive la différence!

The Approach of this Book

In the mastering studio we use the scientific tools of audio engineering to illuminate musical art. So this book constantly integrates art with science. Students may ask why they have to learn all this technical stuff: while you can't get very far without talent, you can get much further with both talent and technical knowledge. In the days of analog processing and analog tape, a practicing audio engineer could get along without a rounded technical education (with the help of a studio maintenance tech), but digital audio requires far more technical knowledge as well as computer skill. A simple "slip of the mouse" can make the difference between a recording that is spatially-compromised and one with a big soundstage. Besides, these days very few recording studios, even the largest, can afford to have a full-time maintenance engineer on staff, so independent engineers need to have far more technical skills than they did in the days of the big studio motherships. A great deal of albums are recorded and mixed in project rooms run by staff with varying degrees of experience and skill. The days of mentorship and education by apprenticeship have largely slipped by. This leaves the project studio engineer and the working independent engineer without many resources to learn their trade. Mastering engineers also need to learn all the techniques that can help mixes produced in substandard mixing environments. All this makes *Mastering Audio* an essential resource for audio engineers, musicians and producers.

Mastering Audio reflects not just the wisdom of over forty years of studio experience, but also my seminar experience teaching audio around the globe to present and future practitioners of the audio art. In this third edition I have completely revised the Chapters and reorganized the book to reflect the mastering engineer's workflow, all the steps from A to Z. We begin with the important concepts that everyone needs to know, building to the more advanced technical ideas. Special terms are introduced for the first time in **boldface** type and will also be found in the Glossary at the tail.

MYTH:
Digital Audio requires less technical skill to use than analog.

Attention Gearheads

This book is designed to help you learn to make informed decisions on your own; how audio equipment works, and what happens when you turn the knobs. Just about every day I get a letter like this one from engineers asking me to bless their particular list of equipment...

> Dear Bob, I always master with a Sis-boom-bah brand compressor and equalizer, then I follow it off with a touch of a Franifras enhancer. On the next pass I use a Caramba tool to maximize the sound and then Whosizats dither before going to CD. Please tell me what you think of my choices?
>
> Sincerely, Gearhead.

I usually reply, politely,

> Dear Gearhead, your equipment list sounds pretty extensive, but much more important is how you use it. For example, some of the gear you describe would be entirely inappropriate for some kinds of music....

As I said, mastering is not about "processing" per se: some masters leave the studio with no mastering processing at all. Perhaps the most essential piece of information we can learn is this aphorism written by master engineer Glenn Meadows.

> *"There is no magic silver bullet. There is no one magic anything that will be 'best' in all situations. The ability of the operator to determine what it is that needs to be done and pick the best combination of tools is more important than what tools are used."*
>
> — Glenn Meadows

Glenn's statement also applies to the amount or setting of each knob or control within our equipment. There is no magic threshold, or EQ setting, or ratio, or preset that will turn ordinary sound into magic. Sonic magic comes from the hard work we put into using our tools (musical magic can only come from the music itself). The truth is that in a typical mastering session, each tool makes only an incremental improvement, and the final result comes from the synergistic totality of the tools working together. In these days of mass-gear-marketing by competitive manufacturers, there is too much emphasis on the glitz, fashion and style of the gear rather than its sound quality and principles of operation. While this book is definitely for gearheads (in the sense that it has lots of glitzy pictures and descriptions of gear designed to produce good sound), serious engineers who want to improve their techniques will also find out how their devices function. Audio principles never go out of style, but gear models fade away.

I have carefully chosen the equipment we discuss as suitable for high-quality audio mastering. Regardless, there are far more models available than I have personally experienced, and their exclusion does not mean they cannot perform a good job.

The theories and background covered here are what I consider to be the minimum necessary to become a competent audio engineer in this digital age. Complex mathematics is not required. There are plenty of good foundational basics for beginners, yet the most experienced digital design engineer will find useful information. This third edition is organized so you can move with confidence from Chapter to Chapter, starting with the basics, and have a good foundation for the advanced concepts that follow. If you are not yet a practicing engineer, it will pay to reread the book after a year in the field: that will reinforce all you've learned in practice.

The Changing Audio World

Greater Emphasis on Singles than Albums

Technological change has a large effect on society. As late as 1995 no one had an inkling that the public would decisively turn from consuming physical music and video product to downloading digital files. In April 2003 the iTunes™ Music Store opened. In less than three years it had sold its one-billionth song! The full effect of this paradigm shift has yet to be evaluated.

The quantity of single downloads moved up from a fraction in 2003 to over 1,200,000 in 2013 while album downloads only increased from a fraction to about 100,000 in the same period, according to John Brownlee at Cult of Mac, who attributes this to iTunes having killed the album by not pricing albums at a sufficient discount compared to singles. This change in public preference from albums to singles has dramatically affected the economics of our industry, trickling down to mastering engineers, who used to produce an album a day and now may produce an album every few days coupled with scattered singles throughout that same week, netting far less earnings per week.

Changes from Downloading to Streaming

According to Alison Wenham:

> By the year 2012, 22% of the independent label group *Beggars Group's* digital revenues came from streaming. The majority of its artists earn more now from track streams than downloads.[1]

These changes lead to challenges to sound quality: Downloads and streaming mean increasing use of coded audio (e.g. AAC, mp3), which is compromised in ways that require us to make adjustments in all our practice. Economics have dictated the increased use of production rooms using nearfield loudspeakers with poor headroom, narrow bandwidth, irregular frequency and polar response — played in small rooms with poor acoustics. As mastering engineers, it is our job to be able to deal with these issues and become mentors to these clients.

Visit www.digido.com/media/links.html for all links mentioned in this book.

The Loudness Revolution

As Bob Ludwig mentioned in the Foreword to this third edition, our audio practices are undergoing a giant shakeup that will be felt for years to come. We call this the *Loudness Revolution*, covered in several new Chapters, as well as related developments in the destination media, from downloads to streaming, from iPods to home theaters to high-end audiophile systems. Those who doubt that the loudness revolution is arriving need only buy a new game, since the game audio initiative began producing audio to the strict R-128 loudness standard in August 2013. Or take a look at iTunes Radio, Spotify, or Pandora, which are all loudness-normalized. Currently they use different targets, which we hope to see reconciled (See Ch. 17) .

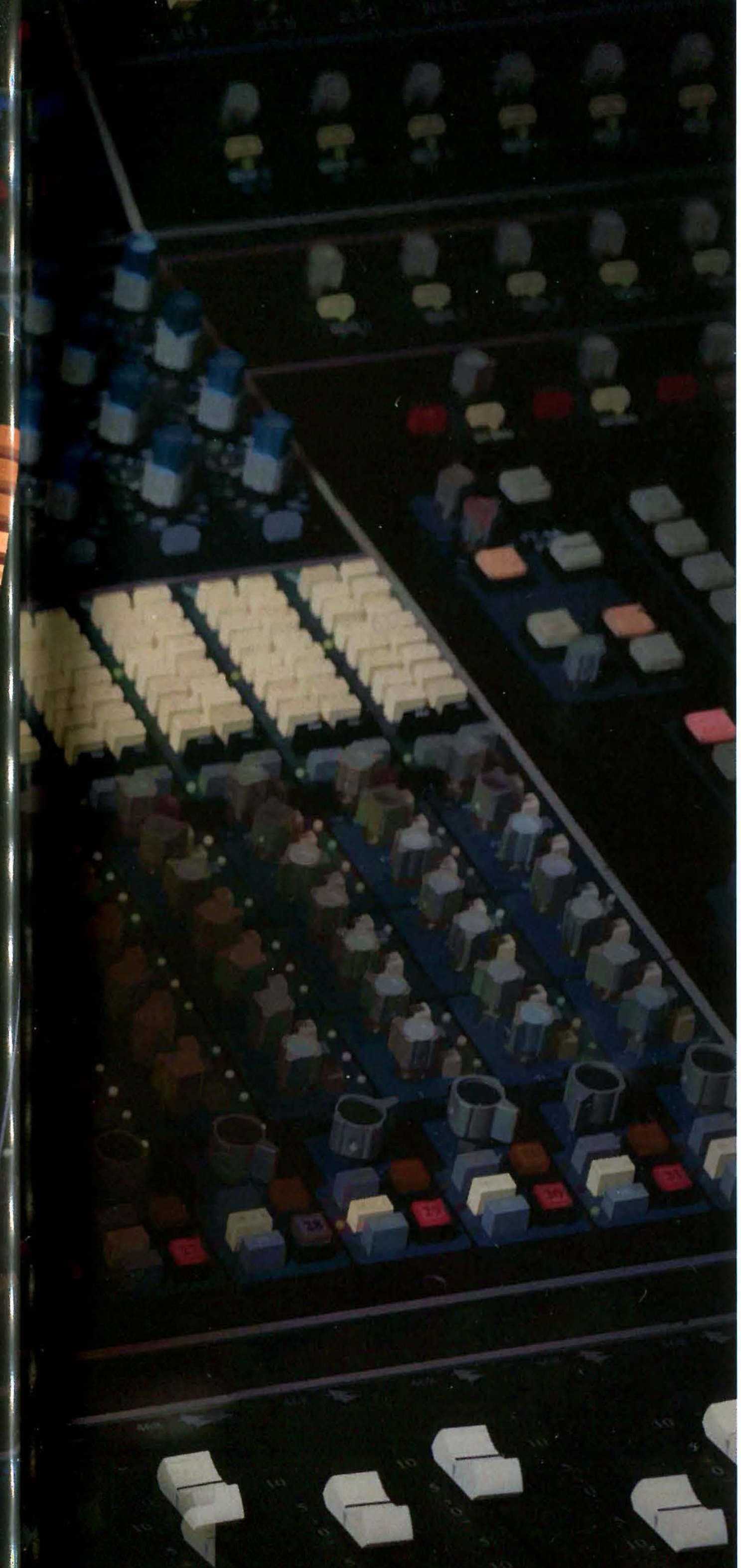
Get Ready to Rumble!

This edition is replete with new, up-to-date information. Just like a well-sequenced record album, these chapters tell a story in a logical, flowing order. We'll begin with an introduction to the mastering world, to mastering tools and procedures, suggest how you can train your ears, and take you step-by-step through a mastering engineer's day.

Let's begin **Mastering Audio!**

---

1 Wenham, Alison (March 2013). *Resolution Magazine*, page 44.



PART I: PREPARATION

"Getting ready is half the job."

—Anon

CHAPTER 1

No Mastering Engineer Is An Island

I. The Evolution of Home Listening

Audio engineers are aware that the consumer's listening experience is constantly evolving. Up to the 1990s, consumers usually listened to their music collections on their home stereo system, and many had sophisticated listening rooms where they spent significant leisure time. But by the mid- to late-1990s, leisure time was more often spent playing video games, watching films on video, social networking, or Internet surfing, either at home (on the computer) or while traveling (on a smartphone). Music listening turned into a casual background activity, using either puny home computer speakers or earbuds while traveling. The consumer's best-sounding music systems have become the ones in their noisy cars! Now they "consume" their music on the run, while exercising or jogging. The novelty of music-on-the-run is mesmerizing, and though mp3 sound improved as Internet bandwidth increased, it generally was played over execrable computer sound systems. It's a sobering thought, but today, most music listening has ceased to be a serious, foreground experience.

From CD, down to MP3, then up to AAC and 96k in the Home

By the year 2000, the CD had lost its attraction for many consumers, who also bypassed the higher resolution SACD and DVD-A, making the first decade of the 21st century the period of the cheap portable digital music player, low-quality mp3 digital downloads, and low-fidelity computer playback. Within the home, the listening experience was deteriorating, but outside, the iPod™ offered a glimmer of hope via its superior DAC and high-headroom audio section (compared to cheaper portable players). The iPod, followed by the iPhone, took portable listening to a high ergonomic and sound quality.

Car sound advanced as well, as auto makers discovered consumers practically lived in the car. Cars became quieter inside and in many cases the systems exceeded the sound quality found in the typical home. By the end of the decade, consumers were playing portable devices in their cars, bringing their large music collections to their primary playback system.

The second decade of the 21st century marks the beginning of the renaissance of sound quality in the home. Apple embraced the superior AAC format over mp3, doubled the bitrate (thus raising the sound quality of songs sold at the iTunes store) and introduced the **Mastered for iTunes** approach, which has been embraced by mastering engineers (covered in my book *iTunes Music*). Apple's iTunes Radio debuted and reportedly has a decent bitrate. Unlike some other streaming services, it does not compress dynamic range and has a very decent loudness-normalization scheme that preserves headroom and sounds much better than terrestrial radio (described in Chapter 16). The loudness race (see Chapters 17-19) is curtailed when loudness normalization is a default in iTunes' file playback. The pure audio Blu-Ray format was devised to succeed the music CD, with greater than 96 kHz sample rate and surround sound. The next revolution in domestic listening was born from a new appliance—the high fidelity music server.[1] This increasingly-popular device lets consumers play high-resolution downloads, Internet radio, music and video files on hi-fi and home theater systems. The device can stand alone, such as the Bryston or Weiss music servers, or be software in a computer, such as the JRiver music server (see facing page) or iTunes modified by Amarra or Pure Music. Each room can have its own wireless connection to the central server, and users play their music collection on demand in any room with a simple remote control. By the end of the second decade, home servers will become the dominant playback mechanism, returning good sound back to the home, as home users will have ripped their CDs to lossless files and begun purchasing high resolution audio files from online vendors. Physical product has become less important to anyone except collectors of fine music (including the vinyl renaissance). It has essentially been replaced by files that can have superior fidelity and that allow instant, convenient access to anything in the consumer's collection.

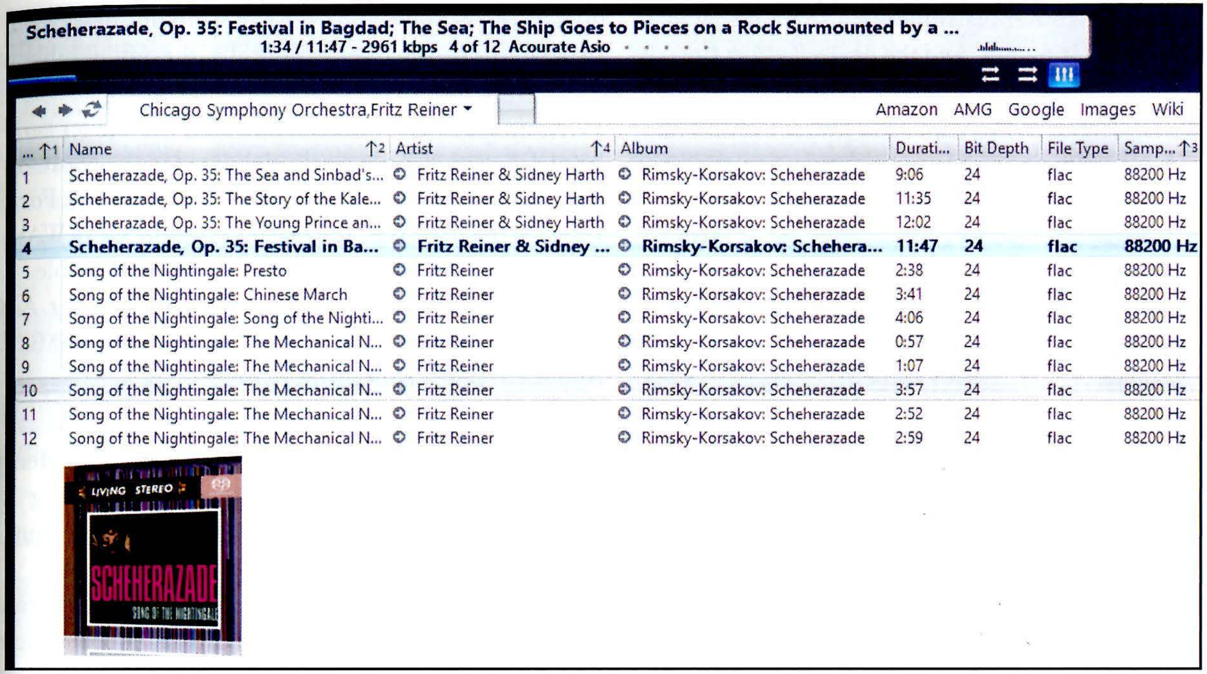
{ Attention to detail: It only takes an extra minute to get it done right, but it takes hours to fix a mistake. }

Mastering engineers know the best way to present audio to the public through this variety of evolving formats and expanding venues. We strive for high-quality audio mastering, and seek to preserve sound quality, reducing it only when the delivery format requires it. And so we urge program producers to create and archive high quality masters, for the future looks brighter than ever before...

II. Technical Steps from Recording to Finished Master

Recording

The computer revolution has given birth to the project studio, where solo artists can produce an entire album in a single room. But the recording process has traditionally been a collaborative one, for music shines when creative people work together. Sound is demonstrably better when music groups are recorded in larger, decent acoustics with instruments that generate acoustical signals (including the electric guitar, bass, etc., which also generate acoustical signals when used with



JRiver Media Center is a computer-based media player which can play high resolution stereo or surround audio or video files from a central server simultaneously in multiple rooms in the house, controllable from computers, smartphones or tablets

amplifiers, which have a sound of their own). After arrangements are written, musicians are hired, and the artists go into the recording studio or on location for the recording to multitrack. For quite some time, the principal medium for multitrack recording has been the computer hard disk, with analog tape reserved for some high-end projects. A primary or secondary hard disk may reside on an Internet-based server, which authorized participants have access to, and which contains a database, musical arrangements, performance tracks, mixes, and later, masters.

Mixing

After the tracking is complete, the producer, artist and mixing engineer produce the mix of each song or section of the work. If you're mixing to stereo, the mix goes to two tracks, but even then it may be divided into several 2-track *stems* in order to produce TV, performance tracks or instrumentals, or to permit some balancing decisions even at the mastering stage, such as the relationship between leads and rhythm (which can be affected by mastering processing). If you're mixing for surround, the mix may go to six or more tracks; and if divided in stems, the surround stems could take up 18 or more tracks!

The biggest difference between mixing and mastering lies in decisions about whether an instrument is too loud or soft in the mix. It's important to make the right decision about this during mixing: it will pay off later. For example if you've mixed the bass instrument too softly, the only way I can fix it during mastering is to raise the level of certain bass frequencies. But this will affect *all* the bass-frequency instruments and the

Replication or Duplication?

Replication is a synonym for *pressing*; the result is a durable molded metalized circle of plastic, sealed under a coat of protective lacquer, which can last for 100 years or more. It is the preferred method for producing 500+ discs. Thousands of pressed CDs can be economically produced in a single day at a CD plant (it takes about 5 seconds for a CD "biscuit" to come out of the molding machine). The master for replication can be a CDR or a DDP file. Replication is highly reliable: once the first replicate has been tested in a special machine, the quality of all the discs is virtually assured.

By contrast, **duplication** means to produce multiple CDR copies of a master CDR using multiple CD writers. A CDR is a chemically-etched medium easily subject to damage from heat and light — though CDR media have gotten much better over time. At 16x write speed, one CDR duplicate can be produced in less than five minutes per slave. Duplication is far less reliable than replication — theoretically each duplicate should be played and verified from beginning to end, since writers can go bad as well as the media. Duplication is more expensive than replication and is advised only for less than 500 copies, though some clients use it for runs of up to 1000 because of the speed of turnaround. Replication plants can be backlogged for weeks.

subharmonics of other instruments. The sound could become boomy or muddy, and the desired effect (to raise the bass instrument) could cause more problems than the supposed "cure" (equalizing the bass frequencies). This is the essence of the problem in miniature: a better mix allows me to make a better master.

Where does mixing end and mastering begin? With the advent of stems, the lines that divide mixing and mastering have become blurred. A good mastering engineer should not concentrate on individual mix instrument levels because this will distract her from the main tasks of integrating the album's songs, and getting the tonality, dynamics and spatiality just right. That is why mixing and mastering should be performed at separate times in studios dedicated for each purpose. Many mastering engineers are uncomfortable with stems if they have never been a mixing engineer. Given a decent mix, I avoid using stems, but if required to use many stems, I reserve time to put on my "mixing head" and tweak the stems in a separate session. Then I start another session and concentrate on the mastering.

Editing and Premastering

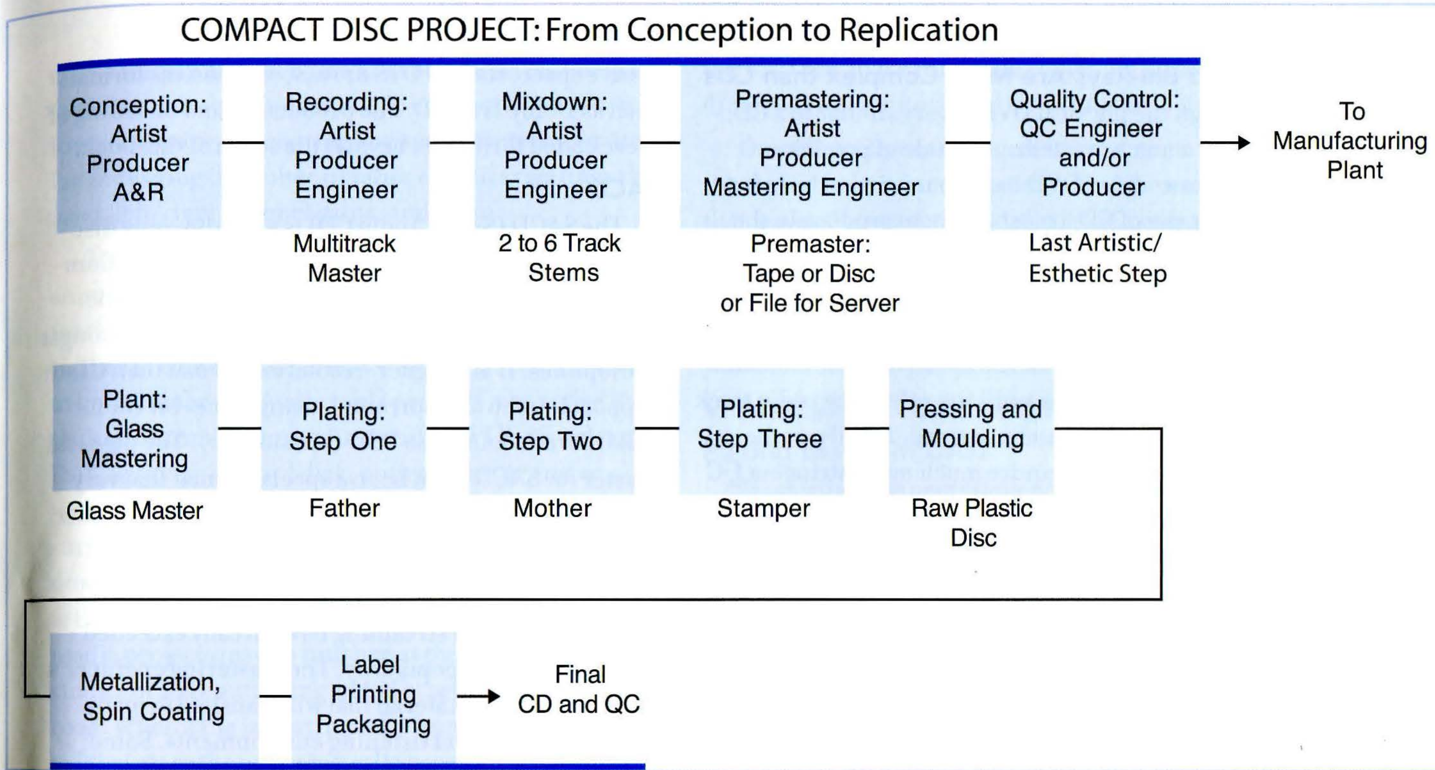
The next step, *editing and sequencing (putting the album in order)*, is usually carried out at the mastering house. Usually sequencing is performed by the mastering engineer, who receives individual songs or segments and puts them in order with spaces or overlaps. Sequencing is followed by *premastering*, which is the proper name of our profession, to distinguish it from the technical mastering that takes place at the plant (though everyone calls us *mastering engineers* for short and we use that terminology throughout this book). Premastering can include the artistic and technical tasks of *dynamics processing*, *leveling*, *equalization*, *noise reduction*, and even some mixing, described in detail in later chapters. Naturally, the output medium of *premastering* is officially called the *premaster*, but we usually label it *master*. For CDs, the master can be a CDR, which is physically delivered to the plant, or a DDP file, which can be delivered electronically. For Internet destinations, the master is one or more WAV (or possibly AIFF files, but WAV is most compatible these days). The conversion to a coded format (e.g. AAC or mp3) is performed by the vendor or distributor (e.g. Apple, CD Baby, Tunecore).

Disc Production in a Nutshell

The compact disc is the most successful high fidelity music medium in history, with a long life beginning with its introduction in 1980. It is still vital to the music industry, although downloads and streaming have rapidly taken over. A finished compact disc master can usually be produced by a mastering engineer in a single day, including the esthetic and technical premastering.

At the Replication Plant

When the premaster is received at the plant, it is used to create the *glass master*. But, technically speaking, glass is not the master. The glass is the carrier for an emulsion that is applied to its surface. At many plants, glass mastering is performed in a class 10 clean room (or better) by engineers wearing white "space suits" (affectionately known as monkey suits). An alternative is to house their LBRs (laser beam recorders) in a self-contained clean room that can be loaded up in the morning by one suited individual, and run all day without intervention, other than to observe the process through a Plexiglas window. The LBR is a multi-million dollar machine that takes the digital information for the master, encodes it[2] to the proper format, then trains an



This figure outlines the major artistic and technical steps in Compact Disc or SACD replication, from the conceptual beginning, through to the finished technical product.

encoded laser beam onto the light sensitive emulsion. The on-off laser pattern generates a series of pits and lands after the emulsion is developed. The coated glass disc is then moved to another clean room, where the emulsion is sputtered with a fine nickel alloy, a process called *metallization*. Next, the disc is put in a vat where an electrical charge is applied, allowing the surface to be plated. This process is called *electroforming*. After plating, the metal piece, now called the *father*, is peeled off the glass. The glass surface is then cleaned and can be recoated and reused for a new "glass master."

The *father* is the inverse of the final disc (pits are lands and vice versa). For small runs, the father can be used directly as a stamper. But for any significant quantity, the father is electroformed to create a *mother* (which is the inverse of the father) from which many stampers can be produced. Each stamper goes into a press, where a clear polycarbonate disc is inserted and molded. Afterwards, the disc is metallized with an aluminum reflective layer (gold can be used in specialty pressings) and coated with a protective lacquer. Finally, a silk-screened or offset label is applied to the top of the disc, which is then packaged with booklets into the disc cases by automated machinery. Every element must be carefully inspected for defects – and the disc itself must meet the proper tests for pit depth and spacing (e.g. jitter and RF output tests). It's an exacting process, but DVDs and Blu-Rays are even more difficult to produce.

DSD and DXD

DSD (Direct Stream Digital) is a digital format that uses delta-sigma modulation instead of pulse-code modulation (PCM). A small group of mastering engineers support DSD and work in various ways. Only a few pure DSD digital processors have been made, so most engineers work in other formats and convert at the end to DSD. Some engineers originate in DSD because they like the sound of the conversion, then convert to high rate PCM by analog conversion or sample rate conversion to retain some of the flavor or warmth of the analog-to-digital converter (ADC) or because they think this produces superior results. Some may originate or edit in DXD (Digital eXtreme Definition), which is 8 times the rate of the CD, 24-bit/352.8 kHz PCM, considered to be the "grandfather format". DXD is editable and processable in a standard workstation, provided that the CPU is up to the task. After processing, engineers downsample and convert to other formats.

I tested a DXD converter and found it to be audibly transparent, indistinguishable from an analog original, so we'll call DXD the reference for purposes

⟶ page 13

DVDs and Blu-Rays Are More Complex than CDs

Although the physical DVD is very similar to a CD, it requires a much greater magnitude of precision. Because a one-sided DVD has about 7 times the information density of CD, it costs more to produce in the creative, technical and manufacturing stages. The creative department has to generate the graphics and menu copy and the plan for interactivity well in advance of the authoring stage; furthermore, all of these elements might be in constant flux until the reference audio track has been firmly edited and mastered. Finally, at the pressing plant, DVDs require much more stringent QC (Quality Control) standards than CDs, especially because of the delicate bonding process for a multi-layer DVD. The Blu-Ray has even more information density and so much complicated capability that all of the above issue are exponentially greater.

III. Specialized Audio Delivery Formats

The Audio-Only Blu-Ray Disc

Most of us are familiar with the high-definition picture and sound afforded by the Blu-Ray Disc, but it is possible to produce a Blu-Ray disc without picture, containing high quality stereo or surround sound at high sample rates. A company called MSM has devised a scheme called *Pure Audio Blu-Ray* that allows the user to select audio formats (e.g. stereo or surround) and navigate to sections via the player's remote control without requiring a video monitor, though track lists can be viewed on a monitor if desired. As CD sales subside, the two remaining major music formats will be digital downloads and Blu-Ray audio. The level of interactivity is far less complex than on a video Blu-Ray and licenses have become cheaper, so the costs are now within the range of independent labels. For surround, Blu-Ray audio, like DVD audio, can be encoded in a lossy format to save space, such as DTS or AC3, or a lossless format, such as Dolby Tru-HD. The production and encoding of these coded formats is beyond the scope of this book.

SACD

The SACD (Super Audio CD) is an audio-only disc format. It was intended to be the successor to the Compact Disc, but did not catch on with the larger public. Still, SACD refuses to die, having found a niche amongst audiophiles. It is a higher-resolution format than CD. It supports stereo and surround, using a one-bit (non-PCM) format known as DSD (see sidebar). The physical master for SACD must be in a special format that very few mastering houses support, so you must send either PCM or DSD-format masters to a specialty house.

Media Files

Downloads and streaming have already exceeded physical media in popularity. The mastering engineer's job is to produce material that will translate to many disparate media and listening environments. Some download sites cater to the audiophile — for example, HDTracks and QoBuz, which sell 24-bit music files with high sample rates.

IV. The Mastering Engineer's Detailed Approach

In mastering, meticulousness and attention to detail are vital. We've always been called upon to keep careful track of a project from the time it arrives until it becomes the final product. Days, weeks, or perhaps years later, if revisions are requested, the client has a reasonable chance of ascertaining which processes were used by consulting with the mastering engineer. At RCA Records, through the 80s, analog tape box labels included "dash numbers" (e.g. -1, -2, -3) for each copy generation, and a card catalog carefully logged the tape's status and which one was the correct *master* to use for LP or cassette duplication. When masters were sent for disc cutting, the cutting engineer inserted a written log indicating the Pultec or other equalizer settings they used, left/right channel gains, and so on.

Today, the situation is far more complicated than simply looking in a tape box for cutting information and marking the box with the generation number. Mastering has become a complex art with many stages that have to be documented. Audio-only projects may arrive in multiple forms: Internet files sent via FTP, digital tapes, DAW sessions on hard disk, optical discs or analog tapes. Projects may be two-channel or multichannel surround; they may arrive as full mixdowns, partial mixdowns (stems), or combinations. The definition of what is the **Master** becomes even more vague, since multimedia projects may be finished at the audio mastering studio, or have authoring added at some studio down the road. When a file is sent digitally from mastering house to label, does the copy magically become the master after it arrives? The answer is: both are masters according to their file names. The safest master is one that is accompanied by an MD5 (a form of checksum) to confirm it has not been changed. But since most label personnel don't understand MD5s, to guarantee that the client receives an exact copy of the master, always zip files before transmission. It becomes useful for the label to retain the zipped file that was transmitted, because a zip file contains an internal CRC (another form of checksum). A file will not unzip if it was corrupted at any time, especially during transmission. It will only unzip if it is intact, and by the nature of zips, it must unzip to exactly the same data as the original source of the zip. Good practice would be for the label to unzip the container with the masters sent by the mastering house, audition the wav files, make reference copies onto another medium, but send the original zip file to the distributor.

One thing has not changed: it is the responsibility of the mastering engineer to ensure that the audio quality that leaves the mastering studio is the same quality that will be represented on the final medium. We must know the project's destination when it leaves our office, and familiarize the producer with what is necessary to preserve the audio quality.[3]

V. Mastering Tools and Procedures

Picking the Right DAW

Mastering engineers depend on their DAWs (Digital Audio Workstations), which must be powerful, reliable, and have very high data (audio) integrity. Sonic Solutions pioneered the mastering DAW and introduced the Source-to-Destination editing model and interactive crossfade editor. To this day, only a few other workstations or software programs have been dedicated to mastering: Audiocube, Pyramix, SADiE, Sequoia, soundBlade (the successor to the original Sonic Solutions), Wavelab and to a lesser extent, Waveburner. When the production is for download-only, it may be possible to adapt other workstations for mastering, but they may not have not all the features and conveniences mentioned below. Convenience translates to speed and efficiency. I probably finish each day an hour earlier because I use a DAW which facilitates the mastering workflow.

Here are some specific advantages of dedicated mastering DAWs:

- Integrated CD track marks and ability to cut and verify master CDRs and DDPs, generate MD5 checksums, insert metadata (e.g. CD text, EAN, ISRC).
- High data integrity; the architecture is designed to be

of discussion. I find other high rates of PCM to be excellent, very close to the DXD original, with 24-bit/96 kHz being the most practical high rate to use at this time for mastering. The losses are very difficult to hear. In a shootout between the 352 kHz original and a 96 kHz reduction or a DSD reduction, I felt that the 96 kHz PCM retained the transients, impact and definition of the DXD original better than the DSD. Conversely, the DSD had a warm sound, which may not be accurate, but is certainly attractive. Let's conclude that when working with such rarified formats, you should pick your poison and use it to its best capabilities: you'll get excellent results. Since 96 kHz is the most practical rate, and most material that comes in for mastering is at 96 kHz or below, I work at 96 kHz most of the time.

"One challenge in mastering is that half the clients complain if their mixes come back sounding radically different, and the other half complain if their mixes DON'T come back sounding radically different."

— Jay Frigoletto

bit-transparent except when performing a calculation.
• High resolution (internal calculation precision).
• Multiple playlists (EDLs) can be open and data can be copied and pasted between them (this saves time).
• Powerful crossfade editor can make "impossible" edits, cutting editing time by over 50% compared to "standard" workstations.
• Integrated cleanup (restoration) facility for declicking and denoising.
• The project supports multiple sample rates, which switch automatically according to which EDL is opened. Track markers created at one sample rate can be managed in a new session of a different rate (this saves a lot of time).
• Clips with different wordlengths and file formats (e.g. WAV and AIFF), interleaved and split files can co-exist in the same EDL.[4] (There is no such thing as a *16-bit or 24-bit* session.)
• Conversion is not needed when importing any audio format, except for sample rate conversion. DAW natively (or via plug-ins) supports export to WAV, AIFF, FLAC, mp3, AAC.
• Integrated dithering. Separate types of dither and dither wordlengths are available on each output (dither is explained in Chapter 15). It is easy to add dither when necessary to output to a master file.
• The waveform shows the effect of a fade and optionally a level shift.
• Nearly instant waveform display as each new file is brought into the project (this saves a lot of time).
• Fades are calculated on the fly, and a crossfade can be any length (this allows easy and smooth creation of long crescendos or decrescendos).
• Some mixing facility, for example, to integrate room tone or to perform segues, to add simple overdubs, or mix stems.
• Object-based processing (available in Sequoia) speeds up the mastering of individual segments or songs, compared to standard "rubberband" automation.

Other criteria appropriate to picking a DAW include software and hardware reliability and economic stability of the company, a well-maintained support structure, the presence of a user group, and ability to submit feedback. All these measures raise the short-term purchase price of a good workstation, but greatly lower the long-term cost of ownership.

Mastering Esthetically

Until about 1967, mastering engineers were "the men in white coats" who cut the records and were not allowed to be creative.[5] Historically, mastering was part of the transfer process, in translating the mix tape so that the record sounded like the mix. Today it is still our goal to present the mix in the best possible way. We should not attempt mixing and mastering at the same time because it defocuses from the goal of mastering.

Our current role falls into one of three basic categories:[6]

1) The mix is done. The mastering engineer may make modest EQ correction, but nothing that would change the mix. Usually the engineers that bring in these types of mixes are very good and have achieved what they want to hear in the mix process.

2) The mix is done, but the producer wants something to happen...

3) The mix ends up not being what the engineer or artist intended, and they are now looking for major changes in mastering.

Every piece of music is unique, and requires an approach that is sympathetic to the needs of that music, the producer and artist. A good mastering engineer is familiar with and comfortable with many styles of music. She knows how acoustic and electric instruments and vocals sound, and she's familiar with the different styles of music recording and mixing that have evolved. In addition, an experienced mastering engineer knows how to take a raw recording destined for duplication, determine what may be lacking, and help make it sound like a polished record. She should also know when to leave a project alone.

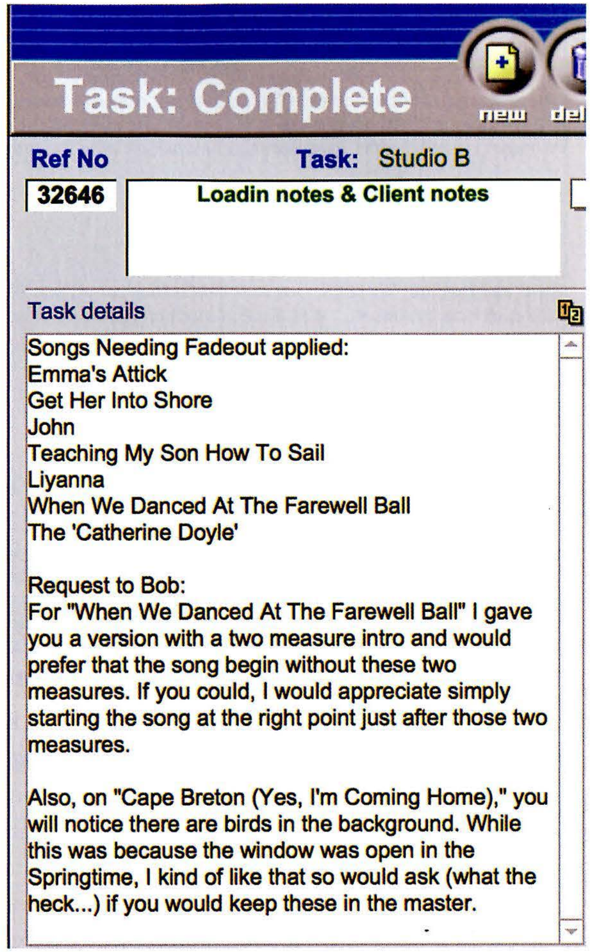
By sympathetically listening to, and working with, the producer, the engineer can produce a master that is a good combination of her ideas and the producer's intentions, better-sounding than if the engineer had simply mastered on her own. The best masters are produced when both the producer and the engineer consult with each other and are willing to experiment and listen to new ideas. As this book progresses, we'll cover esthetic mastering approaches, from the purist to the extreme.

Mastering for the Internet (Streaming and Digital Delivery)

No additional preparation is needed for masters destined for digital download, though level issues should be considered as we will discuss in Chapter 16. Original masters are high-resolution files, which should then be downsampled or data-rate-reduced to the final format. There is an advantage to coding an mp3 or AAC from a 32- or 24-bit file, since it will have subtly better sound quality than from a 16-bit source. My book *iTunes Music* has many helpful hints on the practice of preparing and coding master files for iTunes, and by extension all the download services.

Mastering Without a Producer Present

To help make the mastering process smoother, I suggest a listen/evaluation prior to the session and a discussion of the producer's goals. Then we can master quite comfortably even without the producer or artist physically present. After the mastering session, we'll send a reference disc or WAV files for their approval before cutting the master. Usually by that time, we are enough in sync, needing perhaps to make just minor changes, so there is no need to produce a second reference.



This log is part of a comprehensive database. It contains load-in notes, which serve as a guide for mastering.

Routing Notes

Edit Template   Load Template

A nice "ambient style" club in the box mix which I opened up and made a bit more mysterious and bigger, but the changes are subtle.
Revel mon -10 to help clubs but it will still be low
Sources upsampled to 3296 using Saracon
Sequoia at 96K all locked to Forsell ADC, 24 bit mode 24 pwr-1 dither
Unprocessed duplicate of track 1 placed in the bottom track going out Sequoia 7/8 for source monitoring
Z Sys Route is "DRC Mastering"
All routing in MADI with Totalmix saved as preset #8 in the project folder. Sequoia snapshot "Insert-Loopback" is chosen.

Equilibrium, lowender in the mixer out 3/4 to
Weiss Comp B ME m1 to
Forssell DAC **out** to

ES-8 In: Input full, TH 3/2.5, Output 4.8/4.6 , Manual, Attack 10/10, Rel 2/2, split. SC filter is IN (100 nF).
toning all else bypassed. With th off, output on ADC return is 2.7/2.7 K-14 meter, with thr on it/s 1.5/1.5. Therefore GR is 1.2 dB
to
Bettermaker EQ-230P In: Eprom #2 Memories # 21
to
Dangerous Bax In: 0.5 @ 18 kHz
to

Forssell ADC: Note ADC has +1 dB gain to facility clipping ADC. If not trying to clip, reduce output level of DSB going into the DAC!
Return into Sequoia "from MADAC" to capt 2496 prelim
and out Sequoia to TC E2 doing 1 dB Loud (load a generic scene) to
Return into Sequoia from TC 3/4 to capt 2496 mast and out Sequoia 1/2 to
downsample via Saracon **at -0.2 dB GAIN** to 3244
playback through x dither in a snapshot in Sequoia instance #2

Mastering log containing routing and analog processor settings. Digital processor settings are loaded from a Sysex dump.

VI. Logging and Metadata

At Digital Domain, we keep everything organized with a networked Filemaker Pro database contributed to by the mastering engineer, assistant and office manager. Metadata means "data about data." The database keeps track of the metadata information provided by the record label or artist, such as ISRC codes, album catalog numbers, and barcodes. Each step in the mastering process must be logged to make sure we meet the client's needs on time.

Every mastering engineer has a different approach, but the object of all logging is to be able to reconstruct what was done during the session so as to make revisions or changes easier. As seen here, the log contains load-in notes, processing including monitor gain, engineer's comments, and processor settings.

PQ Lists

The name **PQ** comes from the letter-code abbreviations for the information contained in the subcode of the Compact Disc. The **P** flag is the most primitive flag; it changes state to indicate the beginning of a new track. The **Q** subcode contains information such as timing and program length, copy prohibit or permit, emphasis condition, and ISRC codes. Although a written PQ log is a redundant paper version of what's on the master, responsible replication plants require it so they can see the track titles and the engineer's comments, and verify that the physical master in their possession is the correct one – all part of good QC practice (see PQ list on the next page). If there are any discrepancies, the plant should call the engineer. An exceptional plant will even note noises or over-levels, and ask for engineer's approval before pressing. Still, the major burden for quality control falls upon the mastering house.

EAN and ISRC Codes

The **EAN code** is also called Mode 2 data and is a barcode that contains information about the product, usually the entire record album. This 13-digit barcode is physically printed in the matrix (center) area of a CD and can be encoded in the subcode area by the DAW creating the master. EAN is a superset of the older, 12-digit U.S. UPC code, so older U.S. codes can be

Label Emusica Records, LLC
Title El Rey Del Bajo
Artist Bobby Valentin
Cat # 87731300002

Date 1/7/2006
Source ◉ Analog ○ Digital
Format CD Audio
UPC/EAN 0877313000023

DIGITAL DOMAIN™

931 NSR 434 Suite 1201-168,
Altamonte Springs, FL 32714,
800-344-4361

Mastered by: **Bob Katz**. This master was created on SADiE ver. 5. All levels, fades, & PQ times are client-approved. Please do not alter in any way. Please refer all technical questions to Digital Domain at (407) 831-0233.

Tno	Ind	Start Time	PQ Duration	CD Time	PQ Time
1	01	Hay Craneo			USDBB0600010
	0	00:02:55.67	00:00:02.00	00:00:00.00	00:02:54.67
	1	00:02:57.67	00:03:54.04	00:00:02.00	00:02:56.67
2	02	Arenas Del Desierto			USDBB0600011
	0	00:06:50.66	00:00:04.18	00:03:56.04	00:06:50.71
	1	00:06:55.26	00:04:02.17	00:04:00.22	00:06:55.14
3	03	Guaraguao			USDBB0600012
	0	00:10:57.26	00:00:03.37	00:08:02.39	00:10:57.31
	1	00:11:01.05	00:03:29.44	00:08:06.01	00:11:00.68
4	04	Mi Ritmo Es Bueno			USDBB0600013
	0	00:14:30.32	00:00:03.17	00:11:35.45	00:14:30.37
	1	00:14:33.66	00:05:42.72	00:11:38.62	00:14:33.54
5	05	Codazos			USDBB0600014
	0	00:20:16.46	00:00:03.26	00:17:21.59	00:20:16.51
	1	00:20:20.14	00:03:58.43	00:17:25.10	00:20:20.02
6	06	Cuando Te Vea			USDBB0600015
	0	00:24:18.40	00:00:03.25	00:21:23.53	00:24:18.45
	1	00:24:22.07	00:05:25.39	00:21:27.03	00:24:21.70
7	07	Esperame En El Cielo			USDBB0600016
	0	00:29:47.29	00:00:02.15	00:26:52.42	00:29:47.34
	1	00:29:49.61	00:04:44.39	00:26:54.57	00:29:49.49
8	08	La Vibora			USDBB0600017
	0	00:34:34.08	00:00:04.49	00:31:39.21	00:34:34.13
	1	00:34:38.74	00:04:19.14	00:31:43.70	00:34:38.62
9	09	Aqui No Me Quedo			USDBB0600018
	0	00:38:57.71	00:00:03.19	00:36:03.09	00:38:58.01
	1	00:39:01.32	00:03:49.64	00:36:06.28	00:39:01.20
10	10	Coco Seco			USDBB0600019
	0	00:42:51.04	00:00:03.25	00:39:56.17	00:42:51.09
	1	00:42:54.46	00:03:36.39	00:39:59.42	00:42:54.34
AA	1	00:46:28.73	T00:43:36.06	00:43:36.06	00:46:30.73

PQ Listing showing engineer's comments, track times, ISRC codes and other information

converted to EAN by adding a 0 at the left; the rightmost digit, the checksum, will remain correct. EAN is not required on the CD master, but if I'm given the code, I will enter it and also cross check that the number is a valid barcode. Use a check digit calculator to confirm the value of the check digit at the end of the number, which requires some interesting detective work. If given an 11-digit number by a U.S. company, consider it to be a U.S. code without the checksum. Use the calculator to determine the checksum, and add a 0 at the head. If given a 12-digit number by a U.S. firm, remove the last digit and throw the first 11 digits into the calculator; if the check digit matches theirs, then add a 0 at the head. If given a 12-digit number by a non-U.S.-label, they may still be using U.S. codes, so put it through the same test. If the check digit does not check out, then generate the 13th (check) digit in the calculator. If given the full 13 digits, enter the first 12 into the calculator and verify the check digit. Our database calculates and checks the EAN code automatically.

The **International Standard Recording Code**, provided to record labels by the RIAA, is a unique code for each track on the album. Theoretically this allows automated logging systems to be used at radio stations to track copyright ownership/royalties, but this was only true in the days when radio broadcast music CDs (most radio stations have converted to playing back audio files). The record label provides the codes to be entered for each track. ISRC contains exactly 12 digits; only the digits without any dashes should be entered in the DAW. For example, in the ISRC code: ES-BO1-01-10503, the first two digits are the country code (in this case, ES for España), and the next three digits are the code for the original issuing record label. The next two digits are the year the song was released, and the last five are recording codes assigned to the version of the song itself. That is, Elton John's version of *Your Song* will have a different ISRC code from any cover of the same song. As long as a song is not edited or remastered, it should retain the same ISRC code, even if the rights are sold to another record label; ISRCs are for tracking, and may not reflect the current owner of the title. There is much confusion at labels over ISRC, and I have seen a label who bought another label's assets convert the ISRCs of the tunes over, although this was not necessary. Again, the ISRC does not define the current owner of a song, and it's OK to have the initials of the original label in a song's current ISRC.

CD Plant Read Me Please

We have uploaded a DDP image file to your site for cutting a CD-A.

The title is ______. There are _____ tracks on the CD-A and the total length is _____.

A PQ sheet is included and we DO want CD Text to be encoded on the master. ISRCs and EAN are also encoded in the DDP file.

Please use the MD5 checksum included to verify the integrity of the image.dat file.

****Please note, this file is NOT for cutting a CD-ROM, it is for duplicating audio CDs!!!!****

For all technical questions, please contact .... at.....

EAN and ISRC are also required by vendors providing Internet downloads, but mastering engineers are not responsible for them. ISRC codes currently cannot be put into the metadata of WAV or AAC files (though this is changing even as this book goes to press). They are directly readable from a physical audio CD. Therefore, ISRC cannot currently be used for credit check during Internet streaming of AAC files. Still, vendors usually require the ISRC code for internal tracking and our clients will have to provide it to their vendor.

CD Text[7]

CD Text metadata displays song title, artist, album name, and genre on specially-equipped CD players, most often those found in cars. The term is very misleading for clients who pop their CD reference into iTunes, and expect to see the titles. In fact, iTunes and Windows Media Player get their title and artist data via a database on the Internet and most computer and CD players do not read CD Text from the disc.[8] CD Text is supposed to fit the ISO-8859-1 standard, which includes special characters and accent marks — but only if you want to be daring! In reality, car players around the world produce nonsense characters when they see anything that is not simple ASCII text (no curly quotes, either!), so specify English language, remove special characters, and use the printable characters of the ancient 128-character basic ASCII set, or you will be sorry. There is a character limit of a little over 3000 characters: your DAW should keep track of this automatically.

Although modern mastering DAWs can cut masters with CD text, notify the plant in advance if CD text is incorporated into a project as by default they turn off this facility to prevent spurious characters from being encoded on the pressing. I always send a read-me letter with these kinds of requests (see sidebar).

VII. Media Preparation, Verification, Backups

Mastering Output Formats for CD-DA

While we can accept input (sources) in nearly any format for mastering, only two output formats are suitable for replicating CD-Digital Audio (CD-DA) discs: **CD-DA** (on CDR media), or **DDP** files (Disc Description Protocol image file, sometimes abbreviated DDPi) which can be placed on data disc or uploaded to the plant via FTP.[9] DDP is the more reliable format (nearly foolproof), because it is file-based and files can be compared against a master file for 100% data accuracy. Less reliable is CD-DA, first because it has less robust error correction than a data disc, second because there is no easy way to verify that a CD-DA copy matches the source, and third because clients can play CDR masters (though they shouldn't), and possibly mishandle them or leave fingerprints. Our procedure is to "seal" the CD-DA master in a plastic bag marked "to be opened only by plant personnel".

There are usually 5 files in a DDP fileset (version 2), the most critical being the audio image file *image.DAT*. Auxiliary files *ddpid*, *ddpms*, *cdtext.bin*, and *pqdesc* carry the PQ codes, CD text, version and ancillary information. Some applications use slightly different terms which are also accepted by the plant. Since DDP is file-based, it is customary to include a checksum along with the file set to test for possible corruption during hard disk or Internet copying. The procedure is to calculate an MD5 as soon as the master is made, save that information in a small text file, and pass that file along with the master wherever it is copied. A verification program then compares the MD5 listing against its calculation of the copy, and if they match, then the data must be identical.[10] It also reports if any of the files are missing.

Once a DDP file (or a CDR copy of the file) has been auditioned and approved, it can be copied or transmitted as many times as desired, then checked at the receiving end against the MD5, so every file is a legitimate master, for pressing in different countries if desired. As further protection, the entire DDP fileset should be zipped before transmission.

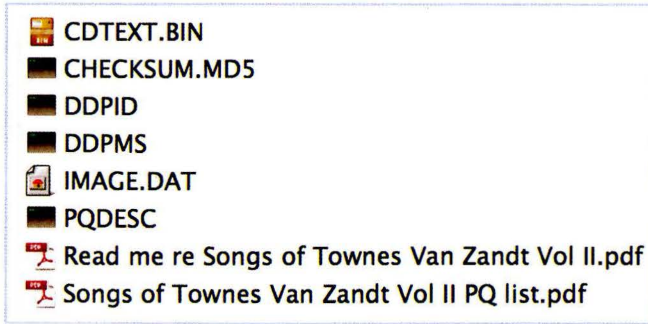
DDP files can also be delivered physically on a data disc, which must be labeled very carefully as the master for producing audio CDs, or else the plant could produce thousands of unplayable coasters with data on them. A CDR rated for 80 minutes of CD audio can hold a DDP image of about 69 minutes, because a data CD holds less data and uses more error-checking. Switch to a DVD-R blank to send an image of a larger audio CD and, again, label it carefully as a master for producing CD Audio discs.

At right is a listing of a DDP fileset ready to be sent for replication, plus MD5, a readme file and a PQ list for the comfort of detail people at the plant.[11]

The master cannot be edited; it must be recorded in one continuous pass, under the control of a computer. Some recording engineers attempt to deliver "masters" on CDRs recorded on a stand-alone CD audio recorder, but this is usually unsatisfactory because of the inaccuracy of the track points, the inability to put separate track end marks (which creates extra-long track times), and the E32 errors introduced every time the recorder stops its laser (breaking "one continuous pass" rule).

References for Clients

Technology is progressing faster than a speeding locomotive, uh, Space Shuttle, err... And we audio engineers are quick to pick up on the changes. In the past few years, the number of physical CDR references



DDP Fileset Listing, plus additional Read Me file, PQ List, and MD5 checksum

I have sent to clients has dropped to almost zero, as has my Fedex bill, thank goodness. Thanks to the speedier Internet and the ingenuity of a single company I can now send electronic references to clients. A company called *Sonoris* has created a method of sending a secure DDP and a playback engine to go with it, as well as the ability for clients to cut their own CDR references. In short, first we make a standard DDP, which is ready for factory replication. Then we run the Sonoris application, which makes a secure copy of this DDP, one that can only be opened and used by our client. The client receives a Sonoris DDP player, which allows him to open this encrypted DDP, play it in his computer, inspect all the metadata (titles, ISRC, etc.), and then cut a CDR reference. The application runs on either Mac or PC. Since this is an exact copy of the DDP master, the client has effectively approved the master. Still, we must QC the DDP master ourselves.

Listening Quality Control

At the end of the project, quality control testing may be performed by a separate engineer, who must have musical/artistic ears, technical prowess,

> *"How many times will you need to QC the masters when delivering a boatload of different formats? Only a fool won't listen to them all."* — Bob Olhsson

MYTH: *An audio load-back/null test shows the integrity of a CD Master.*[16]

and also a lot of common sense: since the project has already been auditioned by the mastering engineer and producer, presumably all the noises were accepted, perhaps even welcomed as "part of the music." But if a single unacceptable tic or noise is discovered anywhere in a master, the entire master has to be remade and listened to/evaluated. There is no shortcut. During the QC listen, which is done with headphones, he may hear noises or problems that were not picked up in the mastering studio. For example, small dropouts on one channel are often masked in loudspeakers. He notes the time of each offending noise, and if it is suspicious, compares it with the original source to see if the problem was introduced during mastering. He would then bring questionable noises to the attention of the mastering engineer. Mastering Engineer Bob Ludwig suggests that headphone listening becomes essential when the number of channels multiplies. Potentially embarrassing noises or glitches hidden in the surround channel when auditioned on loudspeakers become quite audible when that channel is isolated in a pair of headphones. To complicate the situation even further, one consumer might be listening to all channels using surround headphones while others might be hearing stereo reductions (fold-downs). Clearly, a surround master requires much greater attention to detail, and costly time to evaluate, requiring several hours to QC an hour program, including any extra passes necessary to check a fold-down!

"If a single unacceptable tic or noise is discovered anywhere in a master, the full-length master has to be remade and listened to/evaluated. There is no shortcut."

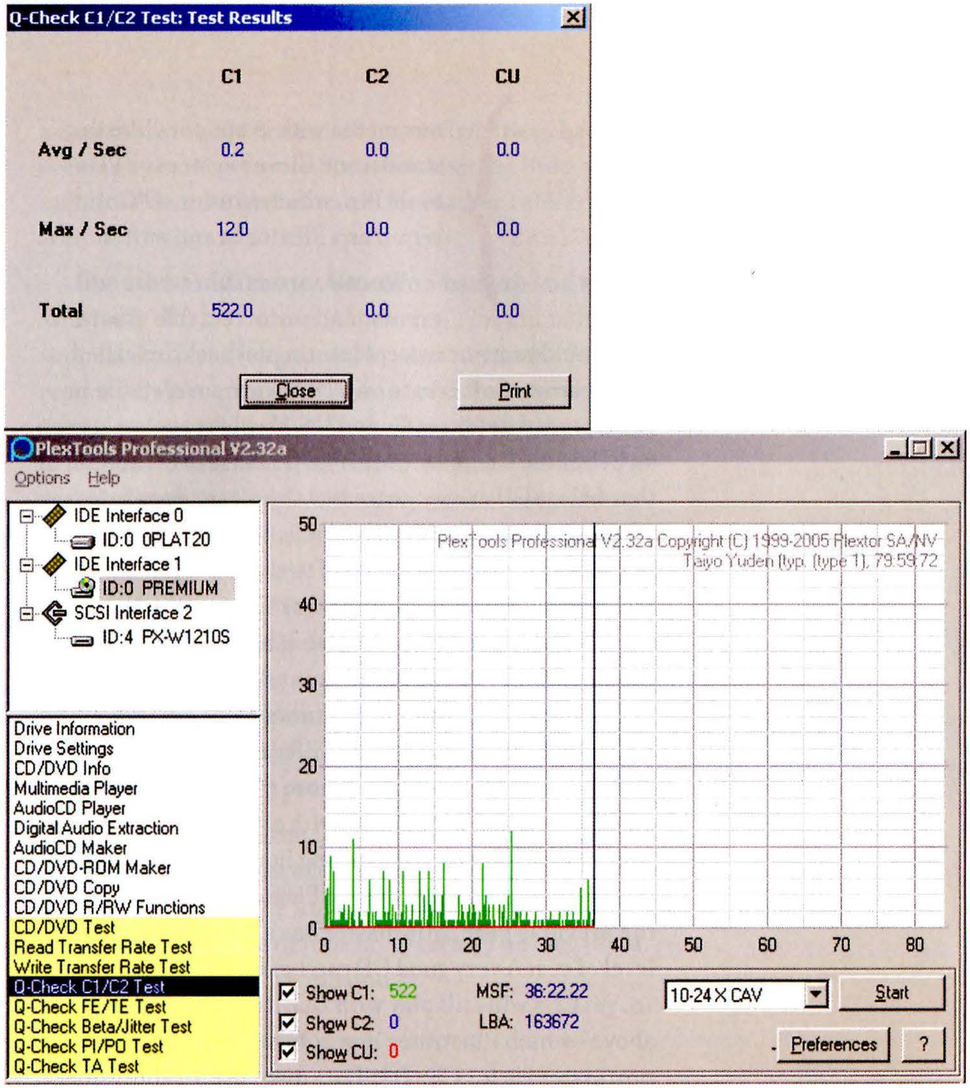
QC includes ensuring that the songs are in the proper place, based on client-supplied lists of the song lengths, lyric sheets, etc., and that the correct master goes out for duplication. We must be especially wary of misidentifying individual CDs of a multiple CD set. Today, with albums in multiple formats — CD, files for download, LP masters, etc., in theory there is no shortcut: each master must be auditioned. But this can be tremendously time-consuming and costly — and often indie clients cannot support the cost of QC'ing every delivery format. An LP master, plus a CD master, plus an iTunes master means 3 hours of QC time! If the budget does not allow a QC of all the masters, then we must inform the client of the risk. Because of the cost of the media, CD or LP masters should always be proofed, but we can be more lenient with files used for upload, since they can be more easily replaced if a problem is found later.

The responsibility for QC must be accepted by someone. I require the client to sign off on every master, so our QC process is for safety, and we usually catch far more problems than the client. The object is to bring the problems below the consumer's radar. There is usually no press proof except when very large quantities are involved. Pressing plants used to have rooms where masters were critically listened to, prior to glass mastering. But now, when the master arrives at the replication plant in physical or electronic form, it will likely be copied at high speed to the factory's central server: no one at all listens during glass mastering. The day has come when the home consumer is the first person to audition the product![12]

Objective Media Verification/Error check

Audio File verification: Since we're in the business of making masters, we tend to be paranoid about data integrity. DDP files and WAV files do not use error-correction algorithms, which underscores the importance of zipping and/or MD5s. Some mastering DAWs have begun to natively support the lossless FLAC format, which does use error-correction and is therefore much more reliable. The first level of protection is that, if any FLAC frame becomes corrupted during transmission, the DAW will know during playback and should notify the user. The second level of protection is that a DAW can confirm that the entire FLAC file is an exact copy of the original audio data. FLAC files are not only safer, they are significantly smaller in size than WAV files, so there is no need to zip a FLAC file for transmission. Apple audio applications do not support FLAC, and can use a lossless format called ALAC. I've read conflicting reports about whether Apple's ALAC format is as reliable as FLAC, but Apple themselves consider ALAC to be robust enough to use for their master storage. Switching to FLAC or ALAC would definitely improve the reliability and speed of procedures.

Optical digital media verification: Optical discs are susceptible to data dropouts that cause errors. This is why all the optical digital audio storage formats utilize error correction algorithms.[13] Uncorrected errors result in glitches, clicks, or mutes. Normally, when playing a disc, we do not know how much error correction is going on. It can sound great, but the disc could be near dying! If the error correction system is working hard, the next time that disc is played, a speck of dust or laser alignment problem, or simply wear and tear, could cause a signal dropout during playback. Our job is to look behind the scenes using specialized measurement tools. Listening without measurement is like having a doctor look at the patient without taking his temperature. So media verification is the internal examination!



CD-DA Report from Plextools Pro. Note the extremely low average C1 (BLER) value of 0.2, since peaks of up to 200 are acceptable by the factory. The peak of 12 in a given second occurs at around 25 minutes as can be seen in the graph.

There is also the issue of *error concealment*, which is the last defense mechanism in digital playback. If there is an uncorrectable error of fairly short duration, instead of muting, the playback machine interpolates between the audio level before and after the error. Short bursts of error concealment can be virtually inaudible or smooth the sound pleasantly, but professionals never use a master medium that is so degraded. So we verify

> *"You're always one generation behind in your auditioning. Unless you proof the copy!"*

our media with evaluators like the standalone Clover System or Plextools Pro, which runs on a PC and requires a Plextor brand writer.

We call correctable errors **soft errors**, and uncorrectable errors that would mute or interpolate on playback are called **hard errors**. Soft errors on CD-DA are correctable in two layers of defense: **C1** and **C2**. Hard errors are known as **CU**. If the C1 correction fails, C2 takes over, and if that fails, a CU error occurs and the player goes into error concealment. If error concealment fails, then the player will mute for a period of time. For replication masters, we do not allow any C2 or CU errors, though we may accept a reference CD that has an occasional C2. And for further comfort, we count the average number of C1 errors per second, also known as **BLER** (block error rate). CD plants permit BLER values up to 200, but our in-house standard is no more than 50, which allows a CDR to age and deteriorate with a margin of safety. Another conservative mastering house accepts BLER up to 100. As we can see from the Plextools error report on the previous page, this master has a remarkable BLER level of 0.2. A very good CD can have a BLER lower than 10, yet CDs will still play with BLERs of 1000 or even above—which illustrates how robust the error correction system is for CD-DA. Data discs use an additional layer of error correction since data cannot be interpolated like audio.[14]

When the CD-DA master reaches the plant, it will be error-tested again and copied to hard disk. However, there is no error testing during the copy to hard disk, which I have already noted is the fundamental difference in reliability between CD-DA and DDP.[15] The plant simply assumes that a CD with a low error rate will transfer dependably in a reader that's in good condition. Millions of CDs have been successfully mastered from CD-DA masters with no problems. But an error test is no substitute for listening back to the master, since when cutting the master, there are many electrical components in the chain after the audition point. You're always one generation behind; if you listen while making a copy, you've only proofed the generation in front of it.

By the way, every computer CD-DA copy is effectively an original, because *soft errors do not accumulate when copying*; the C1 and C2 errors from the source are corrected and the new disc will have its own error count. If, however, the source disc has a hard (uncorrectable) error, a mute or a glitch *will* turn up on a computer copy. Keep in mind that when listening to the disc the error will be interpolated and will not produce an audible glitch. Preparing and error checking DVD-R masters for videos is a specialty not covered in this book.

Backups/Archives/RAIDs

An automated backup program backs up the entire network to a safety hard disk, including audio logs and sequences as well as all the mundane items such as word processing and accounting. Since computer systems and processors are rapidly evolving, we also keep a high-resolution capture file of the master just in case the processor settings can't be recovered.

After a project is finished, we wait until the client has approved the master (usually by listening to a copy of the master). We make an in-house backup of the audio and EDLs on hard disk and then may delete the audio material from our main hard disks. This backup is mostly in case a revision is requested.

We have several servers at our mastering studio. Each one uses a RAID format, which means that the hard disks contain redundant information in case one or more fails. A RAID is not a backup, it is only a protection in case of disk failure. Data is not data unless it's kept in more than one place! In addition, we have further fail-safes, as we cannot afford any down time or loss of data. The first fail-safe is a complete mirror RAID of the audio and document server. This mirror duplicates the source every hour on the half hour. Though it is up to an hour behind, this is an advantage in case someone does something stupid and if he realizes it in time, the last saved copy is on the mirror. We also have a protection from accidental and intentional deletes (it's real easy to hit the delete key on a PC). Any file that was deleted and previously mirrored moves into a special folder on the mirror machine and is kept there for 30 days. This has saved our lives numerous times. In addition to the mirror, we keep an incremental backup of all computers' document files (not audio files) on a server maintained by an application I recommend called *Crashplan*.

I highly recommend a PC-based program called *ViceVersa* for backup, synchronization, replication and comparison. We use ViceVersa for audio file backup, moving files off of the audio server about 30 days after the client approves the project. This is helpful in case the client decides to revisit a project in the future, for example, to reuse a single from a previous project in an upcoming album. ViceVersa is a big improvement over Windows Explorer and the Mac Finder. ViceVersa keeps track of all the files in a group and reports if they were successfully copied, so there are no more missed files (take that Windows, take that Mac!). ViceVersa can use a checksum readback method to confirm that a backup is complete and accurate. As usual, visit the links page mentioned in the Introduction for further reference to archive formats and techniques.

The critical difference between a **backup** and an **archive** is that an archive is copied to a medium that is supposed to last a long time (30 years or more). Some record labels require full backups of the masters and work product, often on DVD-R. What they really intend to mean is "archive". However, if "archive" means "100 years", then hard disk is definitely not an acceptable archive format, nor is any other magnetic or chemical medium. In fact, the best archive of a music CD is the pressing, which could last hundreds of years. Today perhaps the safest long-term archive medium for files is a flash drive, as long as you copy to two of them, and migrate to another medium every 10 years! The next question is whether the equipment will still be around to read the files in ten years?

It is doubtful that the DAW software will be able to read the sessions, especially if plug-ins were used. Make an archive of the final product and don't expect to be able to perform revisions long-term. The idea of full data recovery is truly an illusion.

The last 10% of the job takes 90% of the time.

1
The Slim Devices Squeezebox, introduced November 2003, discontinued circa 2010. The concept lives on in devices like the Bryston BDP-1 or Weiss MAN301.

2
The encoding includes EFM modulation and error correction information. Further references can be found in the links.

3
Here are real horror stories from the trenches: One mastering engineer reported a situation in which another house added the CD-ROM portion to an extended CD, and somehow in the process, changed the audio quality of the audio portion. Never assume that everything will be fine when the master goes out the door, even to the extent of (on critical projects) approving and testing the final product. It is possible to do **null tests** or **bit-for-bit comparisons**, which compare the original audio master against the final pressing, ensuring that the audio data had not been altered after leaving the mastering house. In another situation, a less than reputable plant copied **all** incoming masters using a consumer-based program, which automatically shortens tracks to the end marks, then puts 2-second silent gaps between all the tracks. So the final pressing of a beautifully-engineered live concert sounded like it was edited with an axe!

4
It took Pro Tools many years to add this feature, so it has finally become more useful for mastering.

5
Emerick, Geoff & Massey, Howard (2006) *Here, There and Everywhere: My Life Recording the Music of the Beatles*.

6
Originally suggested by Trevor Sadler, via email and webboard, 2005.

7
Thanks to mastering engineer Jim Rusby for being the original resource on CD Text for the first edition of this book.

8
In iTunes, when a CD is inserted, the Gracenote database is accessed by default (this preference can be turned off). The database counts the number of tracks on the disc, their lengths, and spacing to determine the name of the album, which has caused a few embarrassments over the years (such as when a Christian singer discovered her album came up as hip hop). Currently the best solution is to add or subtract even a frame of space anywhere on the disc until it looks unique to Gracenote. Any iTunes user can submit information on a CD to Gracenote, so it is advisable for the record company to beat the consumers to the punch by uploading data before the CD is released. Content owners can apply to become Gracenote partners, which gives them upload priority and allows them to lock consumers out of potentially disturbing the listing of a title. Some mastering houses have the Gracenote Content Provider application, and can provide the service of uploading correct data to Gracenote. Thanks to Glenn Meadows for helpful tips.

9
The PCM-1630 and DDP on Exabyte tape are obsolete. The master medium used for DVDs is either a DVD-R disc, or DLT (Digital Linear Tape).

10
The slim chance of two different files having the same MD5 is 3 x 10 to the minus 39th power! Therefore, MD5 is extremely reliable and I recommend using it when moving large groups of files from server to server and before erasing the source! On the PC, a shareware program called **Advanced Checksum Verifier** is easy and convenient to manage MD5s. Visit the links page for more suggestions.

11
There are two versions of the DDP protocol. Version 2 can carry CD text information and is now accepted by every major CD plant.

12
Thanks to Mike Collins, *One To One* Magazine, November 2001, and to various discussions on the Mastering Webboard, for inspiring this section.

13
Hard disks generally do not require error-correction, since their error rates are extremely small.

14
Ironically, there is no correlation between a disc's error rate and its readability in a given player, especially delicate players like those in cars. The measured RF signal level is a better measure of a disc's readability; unfortunately, Plextools does not measure RF level.

15
This is the case unless the replication plant adds a custom error-reading interface to the CD-ROM reader which is used to rip the CD-DA to hard disk. The DDP file is copied to the plant's server without any error correction (like any other file copy). After that, plant personnel should run the MD5 utility to confirm this last and most important copy matches the original source file.

16
On the contrary, the null test proves only that there were no uncorrectable errors: it is not a measure of media reliability or error-count. The null test is post the error correction. You could be one bit away from failure and not know it. The next time an error-prone disc plays, there could be an interpolation or a mute if the error count is high. Thanks to Glenn Meadows for pointing out these facts.

{ *Backups? We don't need no ba&*9 u.* }

CHAPTER 2

An Earientation Session

I. Introduction

Ear training is actually mind training, because the appreciation of sound is a learned experience, and the more we experience, the more we learn. Although to our modern ears, Edison's acoustic phonograph gave a crude representation of the original, its first listeners felt that its reproduction was indistinguishable from real life. It is only with each advance in the state of the art of sound reproduction that people become aware of the shortcomings of the previous technology. For example, whenever I work at a very high sample rate, and then return to the "standard" (44.1 kHz) version, the lower rate sounds worse, although after a brief settling-in period, it doesn't sound that bad after all (See Chapter 23).

As we become more sophisticated in our approach to listening, we develop a greater awareness of the subtleties of sonic and musical reproduction. We can also grow to like a particular sound, and each of us has slightly different preferences, which vary over the years. When I was much younger, I liked a little brighter sound, but from about the age of 20, I've tended to prefer a well-balanced sound and recognize when any area of the spectrum is weak or over-present.

A mastering engineer requires the same ear training as a mixing engineer, though the mastering engineer becomes expert in the techniques for improving completed mixes, while the mixing engineer specializes in improving the mix at the level of the individual elements that make up the whole. Ear training can either be a *passive* or a *hands-on* activity. Passive ear training goes on all the time ("what a tinny speaker in that P.A. system"), while active ear-training occurs while your hands are on the controls. Make passive ear training a lifelong activity — it will increase your ability to detect fine sonic differences.

"*Make passive ear training a lifelong activity.*"

Practice being more consciously aware of the sounds around you and identifying their characteristics. Acousticians and classical recording engineers can't help judging the reverberation time of every hall they enter.

Hands-on ear training is the process of learning how to connect technique with the sound you have in your head; like all skills, developing hand-to-ear coordination requires practice. Before working on a piece of music, try to imagine the sound you're trying to achieve, and have a definite sonic goal in mind. Sometimes even if we don't know how we're going to solve a problem, a clear goal can keep us from fumbling.

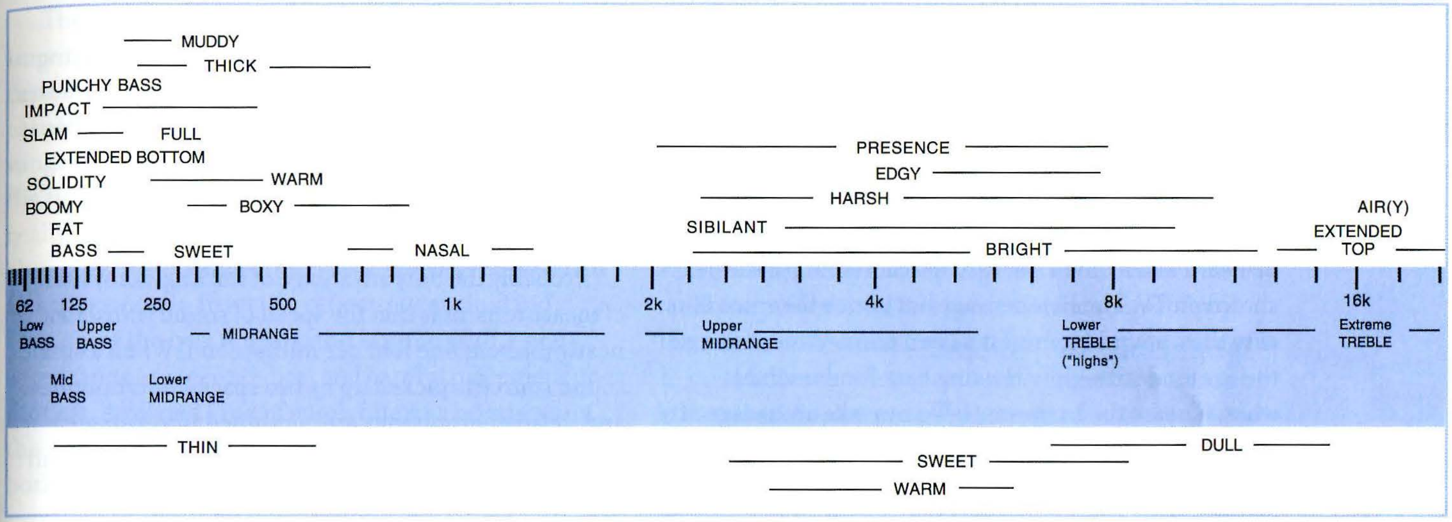
II. Speaking the Language

The classic chart folded into the front cover of this book was hand-drawn in 1941 by E.J. Quinby of room 801 within the depths of Carnegie Hall.[1] We've reproduced it for the benefit of musicians who want to know the *frequency language* of the engineer, and for engineers who want to speak in a musical language. Sometimes we'll say to a client, "I'm boosting the frequencies around middle C," instead of . . . "around 250 Hz." Learn a few of the key equivalents, e.g., 262 Hz represents middle C, 440 is A above middle C, and then remember that an octave is a 2x or 1/2x relationship. For example, 220 Hz is the frequency of A below middle C in the equal-tempered scale. The ranges of the various musical instruments will also clue you to the characteristics of sound equalization — next time you boost at around 225 Hz, think of the low end of the English horn or viola.

Although it helps an engineer to have played an instrument and be able to read music, many successful engineers can do neither, because they have good pitch perception, can count beats, and understand the musical structure (verse, chorus, bridge, etc.).

The chart on the next page is a graphic representation of the subjective terms we use to describe excesses or deficiencies of various frequency ranges. Excess of energy is shown above the bar, and a deficit below. The bar is also divided into eight approximate regions, though there are no standard terms for these divisions: what some people call the *upper bass*, others call the *lower midrange*; what some call the *upper midrange*, others call *lower treble*. Notice that we have more descriptive terms for areas that are boosted as opposed to those that are recessed. This is because the ear hears boosts or resonances more easily than dips or absences.[2]

With an equalizer, the sound can be made *warmer* in two ways: by boosting the range roughly between 200 and 600 Hz; or by dipping the range roughly between 3 and 7 kHz. These two ranges form a yin and yang, which we'll discuss in Chapter 4. Another way to make sound warmer (or its converse, edgier) is to add selective harmonics, as described in Chapter 22. Too much energy, and/or distortion, in the 4 to 7 kHz region can cause an *edgy sound*, especially with high brass instruments. Another common term (not on the chart) is *tinny*, which is probably the same range as *edgy*, but less prominent. Extra energy in the lower midrange, or a strong upper midrange, can add what we call *presence* to a sound, but too much can sound fatiguing or harsh. If the sound is edgy, it can often be made *sweet(er)* by reducing energy in the 2.5 to 8 kHz range. Too much energy in the 300-800 range gives a *boxy* sound; go up another third octave and that excess



is often termed *nasal*. A deficiency in the range from roughly 75 to 600 Hz creates a *thin* sound.

III. Exercises

Ear Training Exercise #1: Learn to Recognize the Frequency Ranges

This is an exercise in the perfection of pitch perception. To have perfect pitch means you can identify each note (or feedback frequency) blindfolded. But this ability is not just a trick: if you learn how to identify frequency ranges by ear, this will greatly improve your equalization technique. I used to practice until I could automatically identify each 1/3 octave range blindfolded, but now my absolute pitch perception is between 1/3 and 1/2 octave, which is about what you need to be fast and efficient at equalizing. Start ear training with pink noise and then move to music, boosting each range of a 1/3 octave graphic equalizer until you can recognize the approximate range. Take a blindfold test, with a friend boosting EQ faders randomly. Don't be dismayed if at first you're only accurate to about an octave: even this will get you close enough to the range of interest to be able to better "focus" the equalizer.

Now let's see how well your musical training lets you be a better engineer. Check out my video *Bass Frequency Surgery*. Not every engineer has been a musician and not everyone can take advantage of the ability to identify musical notes and find the frequency on the Carnegie Chart. But every engineer should know at least the general frequency ranges and how they relate to each octave on the piano. Even if you don't play, if you recognize the notes of the piano, you can use an equalizer with a built-in keyboard, like the DMG Equilibrium pictured on page 60. This has real-world applications. I recently used the Equilibrium to help identify exactly where a close mike was placed over a real piano keyboard so I could dip in the exact center of the range to alleviate a "clangy" quality from the sound.

Subjective terms we use to describe excess or deficiency of the various frequency ranges.

Ear Training Exercise #2: Learn the Effects of Bandwidth limiting

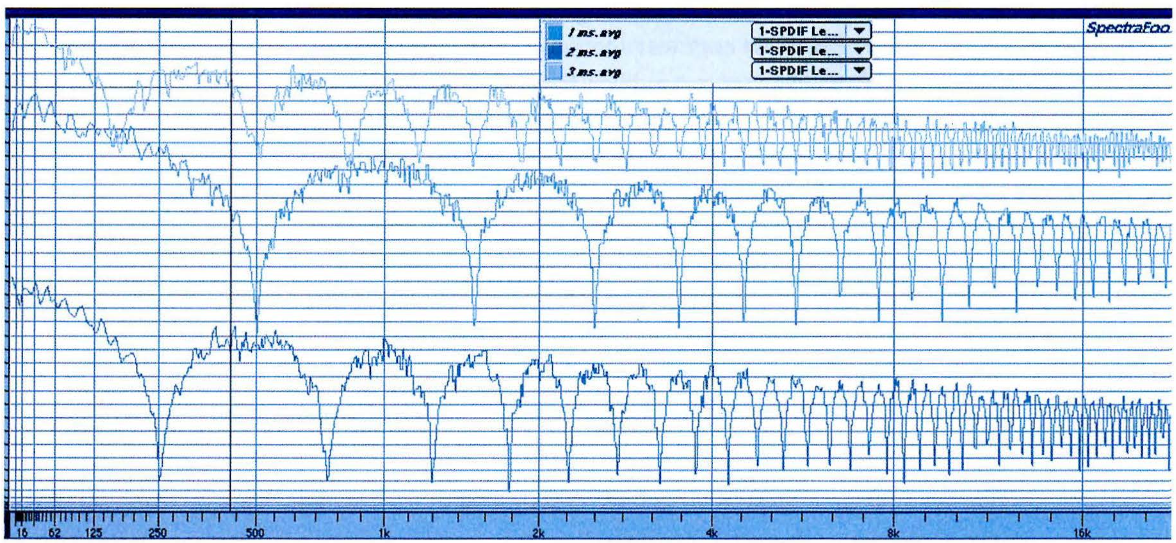
Less-expensive loudspeakers usually have a narrower bandwidth, as do lower-quality media. Train your ears to recognize when a program is either naturally extended or bandwidth-limited. It's surprising how much low- and high-end filtering we can get away with, as can be heard when old films with optical sound tracks are shown on TV. The listener may not notice the voice is very thin-sounding until it's been pointed out, because the ear tends to supply missing bass fundamentals when it hears the harmonics. We can take advantage of this in mastering (e.g., by reducing low frequencies to obtain a higher level), but this is an audible compromise, and the best productions are usually the ones with full bandwidth.

Most musical information is safely tucked away in the midrange – the only frequencies that remain in an analog telephone connection – but a 5 kHz bandwidth takes away the life and clarity of the sound, even if all the informational content is there. Practice learning to identify the effects of bandwidth limiting using high- and-low pass filters on various musical examples. Another way to study the contribution of the low-bass range is to turn subwoofers on and off.

Ear Training Exercise #3: Learn to Identify Comb Filtering

Probably the only advantage of the English system of measurement is that the speed of sound translates neatly to about one foot per millisecond. When a single sound source is picked up by two spaced microphones, and those microphones are combined into a single channel, unwelcome audible comb filtering will result if:

- The gain of each microphone is about the same and the microphones are identical or similar models.
- Relative mike distance from the source is in the critical area from about 1/2 foot (~150 mm) through about 5 feet (~1.5 m). At 5 feet, the more distant mike's signal is lower in level, also reducing the combing effect.



Severe Comb Filtering

Comb filtering can occur anytime a source and its delayed replica are mixed to a single channel; when one source's gain is reduced by at least 10 dB, the comb filtering becomes audibly insignificant. This figure shows the frequency response when source and delay are at equal gain. Vertical divisions are 3 dB. From top to bottom – a delay of 3 ms (about a 3 feet/1m path difference), 1 ms, and 2 ms. In real life, reflections will be diffused and somewhat attenuated, with less obvious effect.

The reflections from a singer's music stand are one important source of this problem, but this cannot be repaired (as some think) simply by adding a piece of carpet, because carpet has no meaningful effect in the range below about 5 kHz, which, as we can see from the figure, is where the major problems occur. The ear really begins to notice comb filtering when the delay is changing, for example, the classic *flanging* effect when an artist sways in front of a reflecting music stand. That's why the best music stand is none at all; open-wire stands are second-best and careful placement does the rest. A related kind of comb filtering occurs when the sound from an instrument reaches the microphone both directly and also via reflections from the floor.

Television and film soundtracks provide excellent laboratory exercises in learning how comb-filtering can mutilate sound, since the proper operation of a lavalier microphone depends on indirect sound, which can include nasty reflections from nearby surfaces. Listen to the TV weather report blindfolded and give your own weather report in response:

"Now she's crossed her hands on her chest, about 3 inches below the lavalier microphone. Now she's turned around to face the blue screen, which is reflecting sound from about 2 feet away. She's sitting down at the anchor desk and you can hear from the hollow dip at 500 Hz that her mike is about a foot above the desk."

Ear Training Exercise #4: The Sound of Great Recordings Well-reproduced Perception of Dynamics, Space and Depth

Because mastering engineers may be called on to work on a wide variety of music, train your ears to recognize good recorded sound in each genre. Start by becoming familiar with great recordings made with purist mike techniques, and with little or no equalization or compression. Learn what wide dynamic range and clear **transients** sound like (see Exercise #15), to more quickly recognize when dynamic range has been limited. Every audio student should be exposed to the sound of real, live, unamplified music, and the sound capabilities of high-headroom, wide-bandwidth, low-distortion, accurate loudspeakers in good rooms.

"Did you know that wearing a hat with a brim puts a notch in your hearing at around 2 kHz?"

The surround medium has the potential to be dynamic. If you have access to a state-of-the-art high-headroom home theatre system, study the superbly dynamic surround soundtracks in lossless coding of motion pictures on Blu-Ray such as: *X-Men First Class*, *Star Trek* (2009), *Ratatouille*, *Nine*, and *Enchanted*. For surround music, try recordings on the 2L label.

Listen to live music: the percussive impact of a real live big band or the clear transients of a classical piano provide a standard that can never be bettered. Compare the depth of a live recital, which can be captured with simple miking techniques, with how much depth is lost when too many close mikes are used.

When comparing a master to the mix, concentrate on one instrument or quality at a time as you switch, and confirm that each stage of the mastering process has made things better, not worse.

Ear Training Exercise #5: The Proximity Effect Game

Most recorded pop vocals have greater lower mid-range and presence than real life. The trick is aided

"The length of silence between two successive plays is proportional to the number of incorrect conclusions."

— Katz's Law

by the recording engineer's use of the proximity effect: increase in bass response when a directional microphone is moved closer to the source. Learn to recognize when a vocalist was recorded too closely, overemphasizing those frequencies (and deemphasizing the contribution of the room acoustics, which can also be detrimental).

Ear Training Exercise #6: The Sound of Overload

When solid-state amplifiers overload, the round part of their output waveform starts to square off. We use the term **clipping** or **clipped** to describe when this overload is severe and the waveform is squared-off. Some amplifiers overload drastically, producing high odd harmonic distortion, others (particularly tube amps) overload more gracefully, which turns them into a form of compressor, fattening sounds when pushed past their linear region. Learn to identify the sound of overload in all its forms: transformers subtly softening transient peaks, analog tape in saturation, overdriven power amplifiers with intermodulation distortion, optical film distortion (as in classic 1930s talkies), etc. Study the saturation on peaks of a classical or pop recording made from analog tape as compared to a modern all-digital recording. Learn the characteristics of each piece of equipment: soon you'll discover some rare digital processors that overload more gently than others. A solo piano recording is one of the most critical listening tests for peak overload distortion. However, a solo snare drum is one of the least critical tests for peak overload distortion (See Chapter 16 for a discussion of the psychoacoustics of clipping).

Ear Training Exercise #7: Identify the Sound Quality of Different Reverb Chambers

Artificial reverb chambers have progressed tremendously over the years. Become familiar with the artifacts of different models of reverbs. Some exhibit extreme *flutter echo*, some sound very flat, while others produce an excellent simulation of depth. We'll learn how they accomplish this in Chapter 10.

Ear Training Exercise #8: The differences between sampled pianos and the real thing

Sampled pianos are sounding better all the time. Sometimes we get fooled! Practice your fine perception until you don't get fooled very often.

Ear Training Exercise #9: Mono, Weak Stereo, Good Stereo

Train your ears to distinguish a good stereo recording from one that has little separation or depth. Distinguishing a mono recording from a stereo is not as simple as looking at the level meters and saying, "Oh, one is moving a little differently", which could be a mono recording with a gain difference between channels. The correlation meter on a true stereo recording should show a variation as the music progresses. An imperfect monitoring environment can give the false impression of stereo information; listen with headphones when in doubt.

Ear Training Exercise #10: Listening Acuity — Identifying Tiny Differences

Make a test master with 0.5 dB difference in equalization of one band. Can you hear the difference in a blind test? 0.5 dB is probably the threshold below which we work on the feeling level, and above which we work on the assurance level. Don't underestimate the importance of audio voodoo; what we believe to be true has a power of its own.[3] However, unless certain, don't be fooled into thinking a difference is truly perceivable. Remember, the longer you wait between listening to one track, and then to another, the less likely you'll be able to accurately compare them. To put it more succinctly:

Katz's Law: The length of silence between two successive plays is proportional to the number of incorrect conclusions.

Ear Training Exercise #11: Habituation

Make a test master that has a frequency response you like. Now make another one that is intentionally 1 dB brighter. Listen to the second master all the way through. Then switch back to the first. You'll be surprised to discover that the first master now sounds too dull! So, which master is the right one? The answer is, probably the first master, because if you're a good mastering engineer, your first choice is probably the right one. One way to ensure you made the right decision is to listen to parts of the brighter master and see if it starts to sound fatiguing over time. Another way is to listen to some of your favorite reference recordings and then play the first master. If it sounds too dull compared to your favorite references, then probably you should turn the highs up.

Another habituation phenomenon is related to the story of the frog and the pot of water. If you drop a frog in a pot of boiling water, he'll jump right out. But if you start him in some warm water and slowly bring it to a boil, he'll boil to death! The human ear reacts similarly to loudness with a phenomenon called *ratcheting*. If you increase the loudness of a recording bit by bit, the ear can habituate to extremes that it might not have if you increased the loudness in one big step. In the EQ example above we might immediately recognize that +3 dB in the high end is too bright, but not necessarily if the EQ change was made in three 1 dB steps. The ear is an excellent relative loudness meter, but not a good absolute one. We discover, during mixing, that a vocal which initially sounded just right above the instruments, is slowly going up and up, because we have lost our objectivity. At that point (or before) we must take a rest. We also need to return our monitor settings to a calibrated monitor gain: this yields more consistent and better-sounding masters (as we will learn in Chapter 19).

Be on the lookout for habituation issues whenever you master — awareness is the first defense. Learn how to be as objective as possible and overcome the ear's natural tendency to habituate.

Ear Training and Monitor Testing Exercise #12: LEDR test

A powerful but simple test for playback system and room acoustics accuracy is called the **LEDR** (Listening Environment Diagnostic Recording) test, invented by acoustician Doug Jones. It's available on test CD JD37 from Chesky Records. If your system cannot pass the LEDR test, then replace loudspeakers, relocate them and/or work on room acoustics. Bookshelf and console-mounted speakers notoriously fail the LEDR test. Learn how nearby reflections destroy the perception of depth and consider moving your speakers and/or treating reflecting surfaces. Details on how to use the LEDR test can be found on the test CD.

Ear Training Exercise #13: MP3 versus CD. 16-bit dither shapes.

Make a 320 kbps mp3 copy of a Compact Disc. Can you hear the difference? Practice until you can identify the differences. Start with a slow-speed mp3 if necessary (e.g. 128 kbps). One of the prime sonic differences will be the sense of depth and space. If your system did not pass the LEDR test, then it will be more difficult to hear the difference between an mp3 and the CD that was its source. An aid to ear training is to listen to the S (side)-signal to expose artifacts of coding (See Chapter 14, page 193).

A related listening test is to compare a 32-bit float or 24-bit master versus a dithered 16-bit result. Dither is explained in Chapter 15. The biggest differences are in depth and space, with secondary differences in tonality. Change dither noise shapes and see if you can identify the shape of dither that best translates to the high resolution original source. I've created a more difficult listening test called *Can You Hear Truncation?* It compares the sound of a truncated or 24-bit dithered result.

Ear Training Exercise #14: The Power of Focus.

Some engineers have made a big deal out of creating listening tests with hidden edits. For example, a hidden edit between flat EQ and a slight EQ change. This test is not only maddeningly difficult but also psychoacoustically invalid — because the ear/brain is not geared to identify changes in a continuous music program. The EQ change will be confused with a musical change at that moment in time. As you search for the hidden edit point, your ear/brain plays tricks on you, depending on the moment in the music where you imagine the edit must have taken place. This doesn't mean the ear can't hear the difference, rather that it is difficult to identify the hidden difference in an edited product. Compare your discrimination ability if instead you play a short piece of music repetitively, once at a flat gain and once with, say, +0.5 dB at 5 kHz. Even if the test is performed blind, your performance will improve! What does this mean in the context of mastering? It means that we can make an EQ change and know it, but it may take some time to know if that change is good or bad. Don't make an instant judgment. Take a minute to listen and see if the change proves wrong when the music gets louder or softer, or if it is a good or a bad long-term change.

Two people can get entirely different impressions from the same playback, because each is focusing on a different aspect — and neither person is wrong!

Another example of the power of focus is our ability to concentrate on one aspect of a production at the expense of others. If you focus on the snare drum in a mix and EQ it without paying attention to the effect on the whole, you'll miss the forest for the trees. Brightening the snare, for example, can easily affect the whole mix and make it sound harsh or fatiguing. Learn to vary your focus, paying attention to specifics, but also to the whole. This is why soloing instruments during mixing is not always a good idea: listen to the context, not just the solo instrument. Of course, experience will improve your performance and judgment. Practice narrowing and expanding your focus while equalizing some already-mixed productions. Ask yourself, "how does this overall EQ affect the guitar, the voice, the keyboard, the bass. Does it help each one of them?"

Ear Training Exercise #15: PLR

The better the headroom of a reproduction system, the more alive and correct it sounds. Observe that a good loudspeaker reveals when a product is over-compressed and how much compression is being used. PLR (Peak to Loudness ratio) is explained in detail in Chapter 16. PLR is the ratio between the average loudness and the momentary peaks of a program — what amounts to its *microdynamics*. Many novice engineers are not attuned to peaks, and it doesn't help if their monitors are already compressing the music. So this exercise is a test of monitors, ears, room and brain. Until you learn to identify when peaks have been softened or removed, please do not attempt to reduce them, because once the peak clarity is lost, it cannot be restored. It's possible to bring the peaks back somewhat, but never as effectively as when left alone in the first place. And of course, it may be esthetically desirable to a given piece of music to soften, or even remove, transients or peak information. PLR is most often applicable to material with percussion, but it also applies to pizzicato strings, guitar plucks, and the short-term movement of non-percussive instruments. For this exercise, I do not expect you to have a grasp of how any meter works or even how a compressor works. This will be explained in Chapters 5-7. The object is to train your ears to identify the sonic effects of transient reduction. Learn to identify when transient peaks have been shortened or cut off, versus when they have been left alone to express themselves.

As an exercise, take a piece of naturally-recorded music that has a prominent snare drum, and pass it through a compressor with a variable attack from 1 ms up to hundreds of ms. Set a ratio of 2:1 to start, a release time of 250 ms, and a threshold adjusted (each time) to produce 2 dB of gain reduction. Adjust the attack time of the compressor, and listen to the effects of the loss of transients. Observe how the attack time affects the definition and even the partial loudness of the snare drum in the mix. See if you can distinguish between attack time of 1 ms versus 150. Then slowly work the attack down until you can identify the difference between 1 ms and 50 ms, then 1 ms and 10 ms, if you can. The smaller the difference, the harder the test gets, and I might have a hard time passing the 10 ms test!

Ear Training Exercise #16: Loudness differences

A good loudspeaker with excellent headroom and definition will let you hear subtle differences in loudness. When I ask my wife to "turn down the TV a little," I usually mean "1 dB, please," while the average listener may mean 6 dB. Practice adjusting a monitor control in small increments until you can tell when the monitor has been reduced by only 1 dB. This is not an easy test, because the music is constantly changing! The more compressed the dynamic range of the music, the easier it is to hear loudness differences.

Now see if louder really does sound better. Compare the sound of two files that are 1 dB apart in gain, but otherwise identical. Which one appears to have more depth and definition? What does that tell you about the importance of loudness consistency in mastering practice? An even more difficult test is to compare two files that are only 0.1 dB apart. This produces a subtle quality difference, not a loudness difference. It is totally possible to tell them apart if your ears are trained.

Next, train your ears to detect improper loudness differences at musical edits. Make an edit at the end of the verse just before the chorus begins. Try to make it your goal to detect a gain difference of only 1/2 dB between verse and chorus. This is not an easy test, because the edit is between two different musical moments. Actually, it's fortunate that our ears are tolerant of edits made between different musical moments in a piece. Still, I go to movies and I notice edits: it's a blessing and a curse! This skill is important to develop not just for editing, but also for learning how to judge the dynamic structure of a song you're mastering.

Ear Training Exercise #17: The Yin and Yang effect

I will explain the yin and yang effect in the EQ discussion in Chapter 4. For this exercise, simply boost the bass a little bit, and observe that the treble seems to get duller — and vice versa. Then, after reading Chapter 4, practice some of the other yin and yang effects and learn why the smile EQ is one of the most popular curves.

Ear Training Exercise #18: The McGurk Effect

Our eyes have a tremendous effect on what we think we hear. Visit http://tinyurl.com/bobkatz2. Expectation has a lot to do with this effect: we all have been fooled by turning an EQ all the way up, hearing the effect, and then discovering that it was in bypass all the time. So don't feel too bad the first time this happens to you. As I mentioned, since music is constantly changing, we do not have a constant reference and so may believe that the equalizer was moved, when it was simply a change in the music. To guard against this, do a lot of your listening with your eyes closed, guard against the power of expectation, take a longer time to reach your conclusions, and cross-check by looking at the bypass switch! A related effect is how the power of peer pressure can actually change perception, or at least your focus, so do your critical listening and reach your critical decisions alone.

Ear Training Exercise #19: Conquer All Your Psychoacoustic Enemies

How do good film mixing engineers produce a film with a consistent sonic center of gravity — despite the fact that a film is mixed in isolated sections that are later spliced together? How do music mixing engineers produce a great recording that comes to a big, beautiful, but undistorted climax yet begins with subtle, soft sounds? How does a mastering engineer produce a dynamic album that intermixes ballads with rockers and integrates **crescendi** with **decrescendi**? The answer is twofold: learn to conquer our psychoacoustic enemies, either through long experience, or serious practice, and master the theory and practice of psychoacoustics, some of which I outline here.

We've learned about (and have hopefully begun to conquer) our psychoacoustic enemy called *habituation* and its cousin *ratcheting*. Those enemies alone are good explanations for the debilitating loudness race. But the real obstacle to good-sounding, consistent mixing and mastering is the nonlinearity of the ear regarding loudness increases (or EQ boosts) versus loudness decreases (or EQ cuts). Psychoacoustician Jim Johnston explains that the nonlinear ear reacts more strongly to a level increase than to an equal level decrease. In other words, the central nervous system thinks that the loudness increase of a level boost is perceptually greater than the loudness decrease of a level drop! Also, momentary *peaks* of a certain measured amount produce a greater loudness change to the ear than momentary *dips* of the same measured amount. Imagine that you are in a bathtub filling with water from a big faucet, but it can only empty from a very small, slow drain. Your job is to keep that bathtub from overflowing. This means you have to periodically take time to turn off the faucet and empty the water. In mixing, this translates to conquering the art of **subtractive mixing**, where we learn how to drop a fader to take some sound away, instead of raising a fader. Subtractive mixing is a very important learned skill, because the nonlinearity of the ear encourages us to raise faders — but it takes a lot more conscious work to bring things down. This also applies to mastering: it takes a lot more work to create a **decrescendo** than a **crescendo**, and we must intentionally bring those faders down more than we had raised them or the total level is going to get out of hand very quickly: all because of three psychoacoustic enemies! One cure is to take regular rests during a mixing or mastering session, say, 10 minutes every hour. When we return from the rest, we will react much more strongly to the loud passage and find it easier to take the level down when a soft passage is required. The nonlinearity of the ear also applies to partial loudness of EQ bands, as we discover that a 1 dB EQ boost at 3 kHz makes the perceived sound somewhat louder than a 1 dB dip at 3 kHz makes it softer!

Ear Training Exercise #20: Unmasking the Mix

Film mixing expert Walter Murch has postulated that the ear can identify only about 3 simultaneous sound elements at any one time in a mix, due to masking. He's described techniques for conquering this issue, as there are obviously many elements present in a music mix, and we must learn to help the listener recognize their ebb and flow. I urge you to study Murch's articles (as always, visit the links page at digido.com described in the Introduction) and learn how to apply his techniques for unmasking background sounds in your mixing practice. Even though the mastering engineer is not usually mixing elements, he should learn these techniques, since to some extent they can be applied during mastering—for example, by using selective equalization to bring out a background instrument and attract the listener's attention, then pulling back on that EQ to avoid continuous listening fatigue throughout the song.

Ear Training Exercise #21: Musical Instrument Identification

Can you distinguish:

- The sounds of soprano, alto, tenor and baritone saxophones
- The difference between trumpet and flugelhorn
- Upright versus "Fender" bass
- In a string quartet, the difference between the first and second violins, viola and celli
- In a large orchestra, the difference between oboe, bassoon and English horn
- Piano recordings: Can you tell when the unisons are perfect or when they are beating?
- Extra credit: Curved soprano sax vs straight soprano

This is just the start of a healthy musical education that I suggest should accompany your audio education. Don't forget the non-western instruments, Chinese, Indian, Balinese... However, don't be afraid to admit to a producer that you are not familiar with a particular instrument: it's just that some of them have unique sounds that may affect how you master, e.g., is it a good or bad thing to warm up a soprano sax? Another thing to watch out for is that each instrumentalist has his own tonal approach: some sax players like to hear their instruments warm, and some brighter or clearer. After a while you can recognize a guitarist or bassist by his sound. Don't be afraid to ask!

Things to Recognize

Experienced mastering engineers have learned to recognize:

- dropouts (digital mutes and analog types), especially audible in headphones
- space monkeys (twitters and glunges, artifacts of lossy coders)
- skewed analog tape
- compression pumping
- hiss
- different frequency ranges of sibilance
- IM gurgle from bad bias in analog tape deck
- phasing (which sounds like varying comb filtering)

- noise reduction misalignment causing pumping
- intermittent noises (ticks, clicks, pops)
- electrical noises (buzz, hum, hiss)
- phonograph associated noises (tracing/tracking distortion, rolloff, swishes, inner groove distortion, non-fill)[4]

Bad Edits. An experienced mastering engineer should be able to recognize a bad edit where the ambience or the sound is partially cut off, or the sound partially drops out. Practice making edits and bring them to the attention of an experienced engineer for critical analysis. You'll be surprised how edits that you thought were perfect may not pass the scrutiny of an experienced engineer (but be kind enough to tell him where you made the edit).

Wow and Flutter. Wow and flutter are caused by speed variations in recordings, and are no longer a problem with digital recording. But mastering engineers are sometimes called upon to restore older analog recordings. So to enhance perceptual acuity, make a cassette recording of a solo piano, and compare it side by side with a digital recording of the same instrument.

Polarity problems. Learn to recognize when the left channel of a recording is out of polarity with the right. Reverse the polarity of one channel and become familiar with the characteristic sound of the error: thin sound, with a hole in the middle of the image. This will also help you to recognize when some instruments in a mix are out and others are in polarity. A good engineer can recognize polarity problems by the vagueness of the stereo image, even without switching to mono.

Recognize the hum frequencies. Hum at the fundamental of the power line (50 Hz in Europe, 60 Hz in the U.S.) usually means a bad shield, an open mike line, or a ground loop. Hum at the second harmonic (100 Hz in Europe, 120 Hz in the U.S.) usually means a bad power supply filter capacitor. Hum at the third harmonic (150 Hz/180 Hz) can indicate induction into an audio cable from a power transformer or a ground loop between chassis. Buzz is hum with lots of high harmonics extending well into the upper midrange.

In Conclusion

Earientation should be a lifelong activity and no one can become an expert overnight. These exercises will help start the process.

1
I've never visited that room, but it would be an interesting archeological voyage to find E. J. Quinby's lair. Internet references indicate he was a renaissance man, and subway train and music expert.

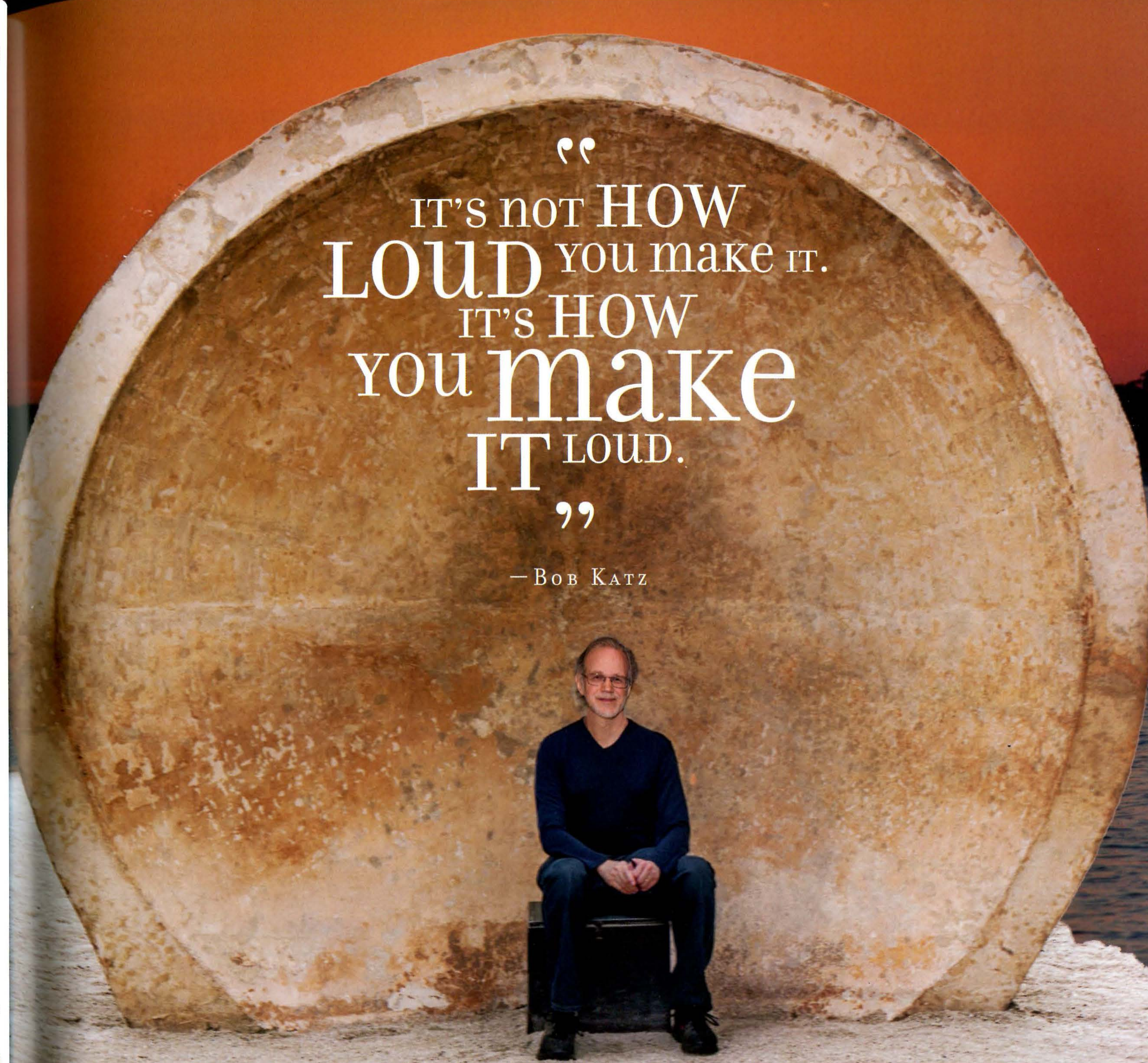
2
Jim Johnston (in correspondence) points out that peaks change the partial loudness more than dips. It's psychoacoustic!

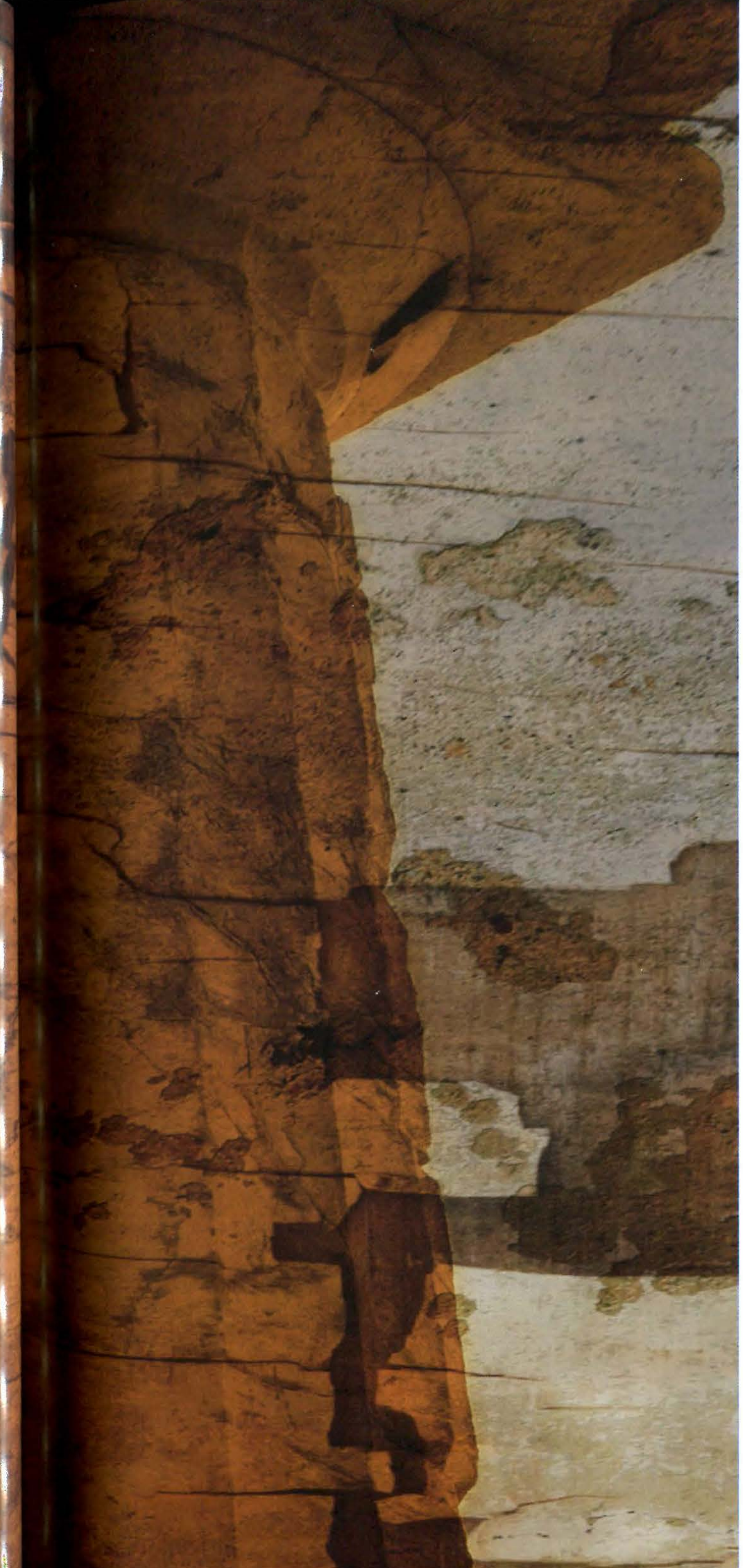
3
Thanks to Andrew Hamilton for that piece of philosophy.

4
Thanks to Jim Rabchuk for this list of items, which should be part of the engineer's listening skill set.

"It's not HOW LOUD you make it. It's HOW you make IT loud."

—Bob Katz





PART II: MASTERING TECHNIQUES

"WE'LL FIX IT IN THE MIX."

—ANON

CHAPTER 3

A Day In The Life Of A Mastering Studio

Introduction

In this chapter, we'll follow the main steps that happen when a music recording arrives at a mastering studio. Our workflow begins with critical auditioning, followed by (if necessary) editing, cleanup, leveling, processing, PQ coding (if it's for a CD), and output to the final medium. This seems straightforward, but you'd be surprised at some of the things we discover!

I. Critical Listening

If there is sufficient time before the deadline, I recommend that the client send a mix of the first-mixed song for a listen/eval even before an album is finished. I audition and give feedback on whether the song is ready for mastering, or if I think further mixing tweaks could make an even better-sounding product. To be done right, critical listening requires years of musical and technical experience, and is done using a very high-resolution audio system. As a general guide, if an instrument's level or tone is off by more than a dB or two, it's probably best to remix, because the better the mix we receive, the better the master we can make. But if all instruments in a certain frequency range suffer from a common tonal issue (usually due to monitoring issues at the mix studio), mastering processing is well-suited to help the situation. Some mix engineers get the bass instrument right, but get the bass drum wrong, often caused by deficiencies in console-mounted nearfield monitors. Mastering engineers have developed some special techniques to improve a bass/kick situation (See Chapter 9), but if it's untenable, we will recommend a remix, or ask the client to send us four stems: bass, bass drum, vocal and the rest of the band. That's why it pays to get that critical audition before proceeding with an album's worth of mixes.

When the raw material (usually files of mixes) arrives, it is loaded into a mastering DAW, then auditioned prior to mastering. The assistant edits and sequences the material if it is for an album, and if there is time, auditions and prepares all of the material, wearing headphones to help catch problems. She listens critically for noises, distortions, and other issues and either brings them to the mastering engineer's attention directly, or notes them in the load-in notes. She may clean up noises using Spectral Editors such as Algorithmix Renovator, Cedar Retouch, Izotope RX or the built-in cleanup/interpolation tools in the Magix Sequoia DAW (See Chapter 8). If there is a problem (such as overload distortion) she will first consult with the mastering engineer, and possibly contact the client. If the client cannot re-output without the distorted peaks, we may attempt to repair them using a declipper (See Chapter 8). She has sequenced the material in the proposed order of the album with spacing that sounds good to her; I will make my own tweaks once I hear it in the context of mastering processing. (See Appendix for suggestions on how to choose the order of songs in an album).

II. Editing

I love editing because a good edit delivers instant gratification. In the mastering studio we usually do not edit together songs from different takes, but we do our share of editing, either cleanup or cutting different mixes together (e.g. vocal up/vocal down). A whole book should be written on digital audio editing techniques, but ultimately the skill of fine editing can only be learned through guided experience: the school of hard knocks, and an apprenticeship. Using sophisticated workstations, we can perform edits that were impossible in the days of analog tape and the razor blade. I have spent 30 hours with a razor blade painstakingly editing a spoken-word version of a novel, a task that can now be done in a single day.

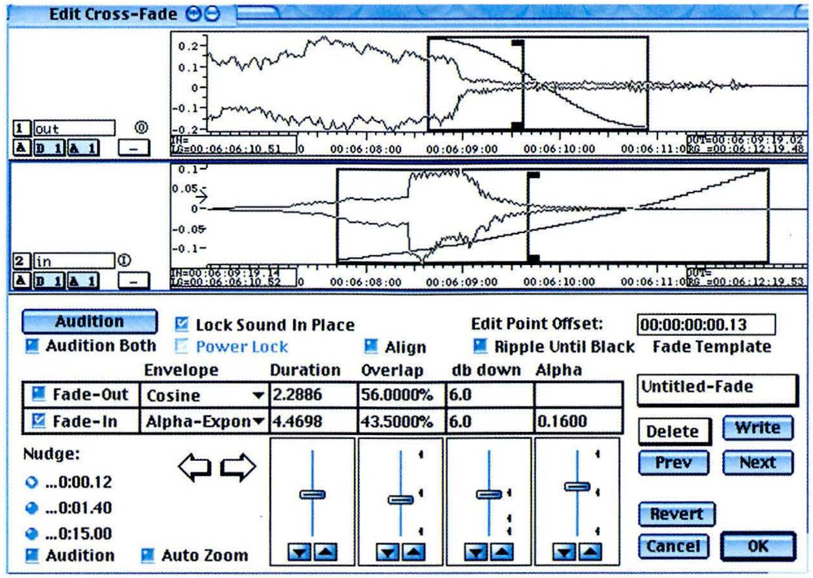
The Tale of the Head and Tail

We have to be aware of the important role played by subtle moments of *natural anticipation:* the human breath before the vocal; the movement of the guitarist's hand before a strum; or the movement of the fingers and keys prior to hearing a piano downbeat. Often it sounds unnatural to cut these off, making the opening appear choked. If the breath is better included, but sounds a bit loud, then a gentle fade-up can produce just the right result. In addition, the ear is jarred by quick sonic changes, between silence and sound, or between two different EQs, until acclimated to the new sound. So even if a mike preamp is fairly quiet, editing from complete silence into the preamp's sound can draw the listener's attention away from the emotional aspects of the music. Sometimes this can be cured by fading in a bit of hiss a few seconds before the edit, then crossfading from the hiss into the preamp noise.

When in doubt, leave the start of the song a little noisy before the downbeat, and send that to the mastering engineer. We are quite skilled at making beginnings that sound natural and have good tools for the purpose. Another cause for that truncated sound is the overuse of samples, which often lack the sound of the air prior to the attack, thus making the attacks sound unnatural. In the hip hop and R&B genres, people have become used to the sound of this jagged editing, which is part of the signature. Regardless, make sure you intend the unnatural truncation effect or it could make the album sound rough and unprofessional.

Follow Fades. Sometimes the tail end of a song contains noise from musicians or equipment, which draws attention to itself in the quiet decay. Interpolation Tools (See Chapter 8) can even remove talking or coughing from a song's decay. Another common solution is called a *follow fade*, which is usually an S-shaped fade to silence placed on top of the decay. A good mastering engineer may spend a minute or more on such a fade to ensure that the tail ambience or reverberation is not cut off as the hiss or noise is brought invisibly to silence. We can take advantage of the fact that noise is masked by signal of the right amplitude, so the follow fade can and should be slightly slower than the natural decay. The delicate decay of a piano chord at the end of a tune should sound natural, even while we manipulate the fadeout to avoid or soften the thump of the release of the pedal. Fine editing can allow us to raise the gain at the tail, after having previously lowered it, in order to hear some inner detail.

Unwhittling the soap. At the tail of the song, fading out is like whittling soap: it can truncate the decay, and normally you can't get back the ambience that you take off. Sometimes we're called upon to make more soap. Sometimes an analog tape may have a lot of echoey or hiss noticeable at the tail of the tune; editing to a quiet digital safety version of the mix, if it exists, can fix the print-through. Or, we can replace the decay with a new artificial tail. Another candidate for an artificial tail is if the musicians or instruments make a distracting noise during the ambient decay; the ambience will sound cheated or cut off if we perform a follow fade to remove the noises. In the figure at right is a fadeout; at the tail you can see the noise made by the musicians. Unfortunately, these noises occurred during the reverberation, so the ambience sounds cut off. The trick is to feed the tail of the music into a high-quality artificial reverb and capture that in the workstation, which you can see in the bottom panel. Since the predelay of the reverb postpones its onset, its position can be adjusted in the DAW's crossfade window which allows us



Adding a tail via a crossfade to artificial reverb.

to carefully shape, time, and level the transition to the reverb in a manner that sounds completely seamless. Thus we have performed the impossible: putting the soap back on the bar!

Bounce with Handles!

One of the most common problems we encounter from inexperienced engineers is audio whose beginning or ending has been cut short. This can result from poor bouncing technique – that is, starting the bounce too close to the beginning of a song, or ending it too close to the ending. When preparing files, play it safe: bounce with handles, adding one or more extra seconds before the first sound begins and after the last sound ends. After the song ends, listen carefully to the full decay in a quiet room to ensure that all the reverberation has ended, and include a one-second or longer handle to be safe! There's plenty of hard disk space to spare, or as we used to say, "tape is cheap."

Ba dEdits

Watch out for extra breaths, which come from multiple punch-ins. Watch too for clicks and other artifacts introduced by edits in your favorite DAW. Noisy edits will be exposed in a good monitoring environment, so listen carefully in a quiet room. Using a Spectral Editor (see Chapter 8) in mastering, we can often isolate breaths and clean up bad edits in an already-mixed song, so now producers expect miracles from mastering engineers with our powerful tools. But it's usually better to fix a problem before the mix is finished. For example, when the vocal track is muted in the relative silence before the instrumental chorus begins, the change to silence draws attention to itself and is difficult to fix during mastering. Instead of muting, do a slow fadeout, or edit into room tone or preamp hiss, fading out after the louder instruments have established themselves enough to mask the hiss.

Another type of bad edit is one in which the reverberation of one section has been cut off by the insertion of a new one. This often happened in analog tape editing because we could not do intricate crossfades with a razor blade. But the error still occurs in classical music recording if an inexperienced producer tells the musicians to begin the retake exactly at the intended edit point, instead of a few bars earlier. The latter approach would not only give the musicians a running start and allow for a better music flow, but would also generate the reverberant decay of the preceding note for the editor to work with. If the producer did not record the reverberation, the ear notices the cutoff of the reverb, which is not masked by the transient attack of the next downbeat. Luckily, when it comes to mastering, we can repair some of these bad edits even if the original takes are not available. The trick is to take apart the original edit and create a short piece of reverb that will overlap the other two elements.

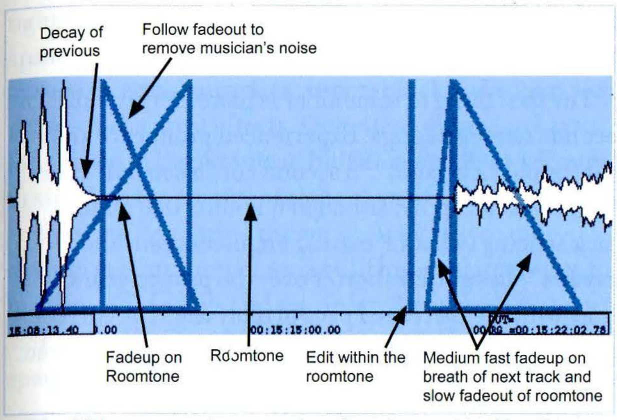
Fadeouts

A good-sounding musical fadeout makes us think the music is still playing; we're tapping our feet even after the sound has ceased. Although we can apply the same S-shape we use for tails, fadeouts are a distinct art in themselves. Typically, a fadeout will start slowly, and then taper off rapidly, mimicking the natural hand movement on a fader. There's nothing more annoying than a fadeout that lingers beyond its artistic optimum. On the other hand, a fadeout should not sound like it fell off a cliff: often in mastering we get material that has to be repaired because the mix engineer dropped the tail of the fade too fast. As we noted earlier, editing is like whittling soap, so I recommend that mix engineers send unfaded material so it can be refined in the mastering. It is difficult to satisfactorily repair a fade that is too fast at the end; sometimes applying a taper on top of the original slope can help.

Adding Room Tone

Room tone is essential between tracks of subtle acoustic and classical music. Recording engineers should bring samples of room tone to an editing session. Room tone is usually not necessary for pop productions, but if a recording gets very soft and you can hear the noise of the room, going sharply to *audio black* can be disconcerting.

Always record room tone in advance as a separate "silent take" with no musicians in the room. If this is not supplied (10 seconds is ok, 20 is ideal), it is almost impossible for us to manufacture a convincing transition and we have to be satisfied with a fade to/from silence. Spectral Editors are handy at creating a glitch-free file of room tone.



Editing room tone in an acoustic work requires considerable artistry. An edit must not call attention to itself.

Live Albums (Concerts)

I love to master live albums: they're fun to edit and assemble, and require fascinating skills to make them feel like continuous performances, even though they are usually recorded and mixed out-of-order and assembled from many different performances and venues. Depending on your skill at assembly, you might want to leave some or all of the editing decisions till the mastering. But in any case, a good recording or mixing engineer will collect a *lot* of room tone samples from between the songs and deliver these as so-called *wild material*:

- loud cheering and applause, several different takes
- audience murmurs at different levels: excited, happy, relaxed, contemplative, quiet, with or without light claps or shouts.
- different kinds of room sound appropriate to transitions between quiet or loud tunes – without the announcer or band voiceover, unless it applies to one specific song.
- quiet room tone with no obvious intermittent audience sounds (or the listener might recognize that distinct shout from the balcony the second time he hears it).
- applause endings – these are precious because it's difficult to find an ending clap that is not interrupted by an announcer or band member or some identifiable sound like tuning that's only good to hear once. Finding a single ending clap that is in the clear (even one second in the clear after the clap) can save an entire album from the deadly "let's fade out here" disease.

Depending on your ability to make an album sound continuous, you may want to deliver either individual songs or a whole sequence. Individual songs are best terminated with ending applause that ends clean with an ending clap or a few words from the band followed by clean room tone. A whole sequence is useful delivered in checkerboard fashion (alternating material on two different tracks): two stereo WAV files, the first containing songs 1,3,5, the second songs 2,4,6, etc., both files in sync, each file containing a hole of the correct length where the other file would be playing music. The hole can contain some room tone, delicately reduced in level after the next song begins, then to silence for most of the rest of the song as the music will mask the removal of the room tone. Or sporadic wild audience cheers can be inserted underneath the other file's music performance. Include notes as to which venue each song was recorded in, and any problems to look out for. We will then take this assembly and adjust levels, eq, or tonality to complete the illusion of an exciting, continuous concert performance.

Applause and room tone. To edit applause requires a familiarity with natural applause, but real-life applause is almost never as short as edited album applause (typically 15 to 30 seconds), and real-life artists have to stop to tune their instruments. The object is to capture just the essence of the concert so that the home listener is never bored on replay. Joining applause and ambience from different performances exercises the power of the workstation's crossfades and the editor's ears. There must always be some degree of room tone (audience ambience) between numbers, but it cannot be a continuous loop because the audience noise at the end of a loud performance is louder and more enthusiastic than after a quiet one. The transition from a loud to a quiet number has to sound natural and will not sound realistic if you simply drop the ambience level: it must be done with a careful crossfade from a boisterous audience into a contemplative one. My approach is to do the major ambience cutting on one stereo (or surround) track, and wherever it needs transitional help, mix in a bed of compensating ambience on another track pair. I once put an audience loop under the only studio cut on a live album, and to this day no one has been able to figure out which track is the ringer!

III. Spacing The Album

Although we are in an era of digital downloads, with its emphasis on singles and the shorter attention span of today's listeners, the record album is still an important music medium. *Sergeant Pepper* is often cited as a seminal rock and roll *concept album*, i.e., an elaborately-designed album organized around a central theme that makes the music more than a simple collection of songs. This started a trend that many assume has more or less died. Is the concept album really dead? Not for me; I treat *every* album that comes for mastering as a *concept album*, even if it doesn't have a fancy theme, artwork or gatefold. Processing, song spacing and leveling contribute greatly to the listener's emotional response and overall enjoyment of an album.

The first thing to remember is to never count the seconds between songs. Experienced producers know that the old "4 second" "3 second" or "2 second" rule really does not apply, although it is clear that album track spacing (what LP cutting engineers call "the spreads") has gotten shorter over the past 50 years, along with the increased pace of daily life. The correct space between songs can never be accurately measured, for different people start counting at different times depending on when they think the decay is over. Counting from the beginning of true silence, the computer may objectively say that a space is only 1 second, but the ear may think it's closer to 2.5. So don't count – just listen. As a general rule, the space between two fast songs is usually short, between a fast and a slow song is medium length, and between a slow and a fast song is usually long. After a fadeout, the space is usually very short, because the listener in a noisy room or car doesn't notice the tail of a fadeout. Often we have to shorten fadeouts and make segues[1] or the space will seem overextended, especially in the noisy car. Perception of appropriate spacing can depend on a producer's mood or even the time of day. If you space an album in the morning when you're relaxed, it almost always sounds more leisurely than one which has been paced in the afternoon, when our hearts beat faster. Be aware of these external factors when spacing an album.

The overall *pace* of an album is affected by intertrack spacing. We probably want the first set to be exciting, so we control the pace using shorter spaces within the first set, and slightly longer spaces thereafter. Interestingly, our sense of timing is relative; if we begin with very tight spaces, then revert to "normal" spaces, the normal spaces seem too long. Manipulate spaces to produce special effects – surprises, super-quick and super-long pauses make great effects. One client wanted to have a long space in the middle of his CD, about 8-10 seconds, to simulate the change of sides of an LP. This sounded crazy to me, but I tried the super-long space, and despite my intuition, it worked! This was largely due to his choices of songs and the order. The set which began side two had a significantly different feel, and the long space helped to set it off, like a concert intermission.

Spacing Rehearsals. Some people think that it's sufficient to play the last 30 seconds of a tune in order to judge the space before the next. But if you play the whole exciting song, you will most assuredly need more time to catch your breath before the next one can start. To avoid playing a whole song, we keep this effect in mind; when we play the album through, we'll know if we were successful. One technique for judging a space is to cut it shorter and shorter until it is obviously too short, and then add just the scintilla necessary to make it sound "just right" – especially knowing that it will seem longer in a domestic living room. Another spacing technique is to make the downbeat of the next song be in time with the rhythm of the previous. This can sound very nice if not overused.

Analog tape editing did not facilitate these kinds of changes. It's interesting to note that when an LP master comes in for conversion to CD, the spaces always seem too long. One reason, as mentioned before, is the current quicker pace of modern life. But the other is that vinyl and tape noise acts as a filler. When there's dead silence between tracks, spaces always seem longer. Remove two or more seconds from an LP space and it might feel just fine on CD.

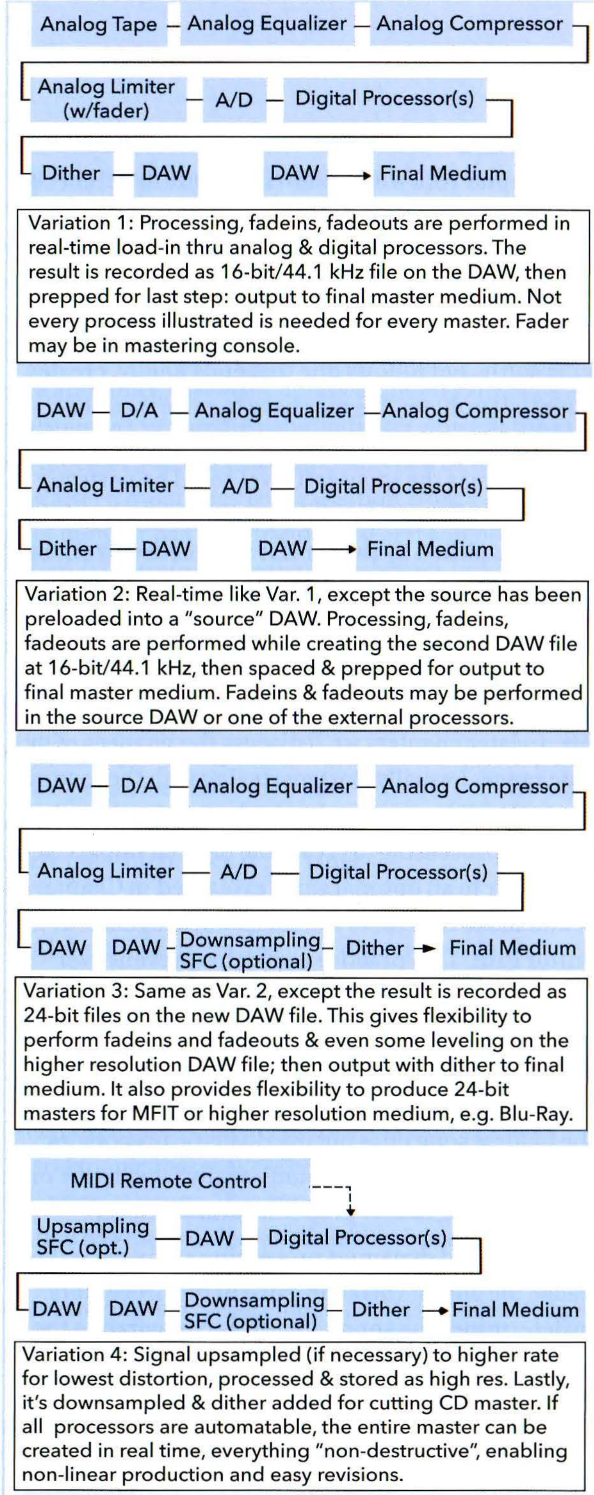
IV. Leveling The Album

In Chapter 16, I will explain the subtle differences between the terms *level, loudness, gain*, and *volume*. For now, let's talk about how we manage the relative levels of the songs in an album.

Context-based leveling. A piece of music that begins softly, but follows one that ends loudly, creates a potential problem. We may have to raise these beginnings because the ear had been acclimated to loud sound, whereas the same soft level could be perfectly acceptable in the middle of the piece. Similarly, a loud attack is amplified by the ear if it is preceded by silence. This is why albums and singles may have to be leveled differently.[2]

The greater a recording's dynamic range, the harder it is to judge "average level": you have to listen in several spots. In later chapters we'll discuss the nuances of dynamic range. I usually start mastering with the loudest song on the album and find its highest point. I then engineer the processing to create the desired impact, set the monitor to an optimal gain (explained in Chapter 19), and make the rest of the songs work together at that monitor gain. This practice also helps prevent overprocessing or overcompression[3] (See Chapters 5-7). During the processing of this loudest song, it's important to ensure an optimal gain structure in the chain of processors. This is the test for the rest of the album, for if the loudest song does not overload the processors, neither will the rest. The album usually falls in line once the loudest song has its proper level and impact.

Leveling and dynamics processing are inseparable, because the output (makeup gain) of the processors also determines the song's loudness compared to the others. A more compressed song may sound louder



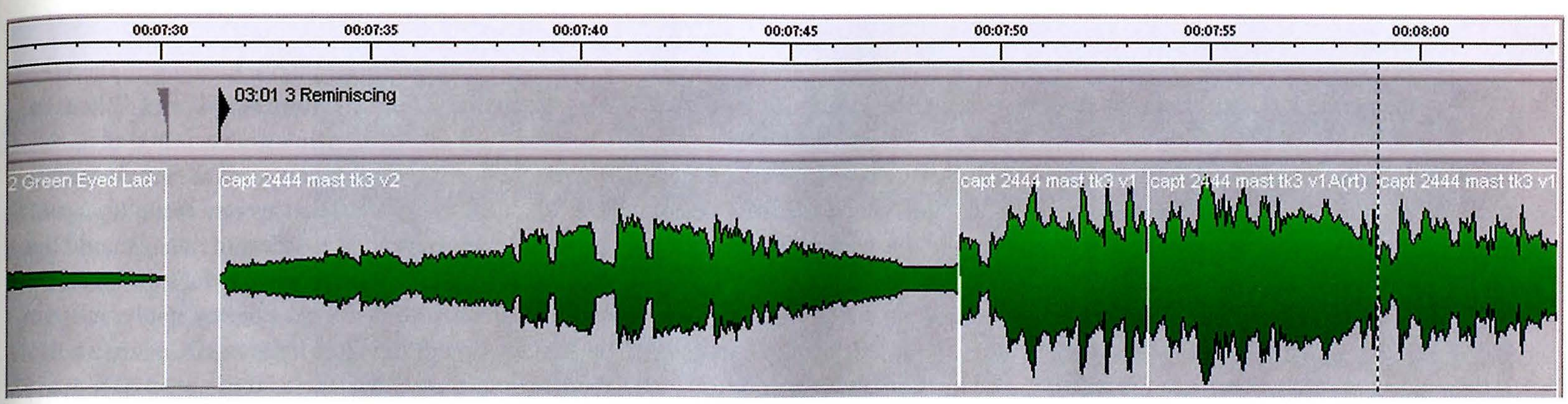
Infinite Variations on a Mastering Theme. Four examples of approaches to audio mastering.

than another, even if its peaks don't hit full scale. If we change the processing, we have also changed its level: it has to be done by ear. After working on the loudest song and saving the settings, I usually go to the first song and proceed in sequence. After I master the second song, I recheck the transition between the first and second. This transition will usually work without any fine-tuning because I've been monitoring at a consistent gain. If one song appears too loud or soft in context, I make a slight adjustment in level until they work together, or sometimes increase the spacing to "clear the ear." So you can see why it's important to have the album in proper order before mastering!

Consider the size of the ensemble. A song with solo vocalist and acoustic guitar should be naturally softer than the full electric band. During the course of mastering I often will turn the ballads down and the rockers up. This is because many mix engineers still follow the practice of maximizing levels for each tune, which was useful in the days of analog tape, but is not necessary with digital recording. This is why I advocate trying to mix by ear, not by meter, using a constant monitor gain or reference. Even though it's important to check the song mix at different levels, returning to the monitor reference will help produce more consistent mixes, both tonally and in the context of an album (See Chapter 19). In other words, try to master the album to some degree during the mixing stage. Have no concern about signal-to-noise ratio when mixing to 24-bit: it's only low if it sounds low!

Everything Louder than Everything Else

After leveling and processing the last song, I review songs one and two, to make sure they still fit well into the context, or if there is a tweak that can further optimize them. Or, I might find that the album has grown



Matched Edits in an EDL combine multiple revisions and save production time.

in amplitude due to ear fatigue and the later songs may need to be lowered. This gives non-linear production a distinct advantage (explained below).

Overzealous leveling practice can produce a *Domino Effect*. Suddenly, the song that used to be the loudest, doesn't sound as loud. Every song can't be the loudest one! If it was loud enough before, the problem may be unintentional escalation. Instead of trying to push the loudest song further, thereby squashing it with the limiter, it may be wise to lower the previous song by even a few tenths of a dB, which will restore the next song's impact by use of contrast.

V. Processing

Now that we have covered the steps of critical listening, editing, spacing, and leveling, we come to processing. Processing is a significant part of our job, although I want to reiterate that mastering is not about processing per se. Experienced and skilled clients do not expect us to change the sound of their hard work, just polish what already sounds good. With that in mind, we'll spend the rest of this chapter showing how we go about processing.

First we listen briefly to the songs and try to decide if analog or digital processing is going to be best, or some hybrid of the two. Many engineers work with DAWs in very much the same way we worked before there were any DAWs.[4] First, we take the source for each tune (e.g. analog tape or digital file), and process one song at a time. If that source is analog tape, we may master the tape in real-time during load-in with analog processors. If that source is digital and if we desire to use analog processing, we send it to a high-quality DAC, pass it through one or more analog audio processors and possibly control the level, EQ, or fade via a customized analog mastering console. The signal is then passed to a high quality ADC, optionally through various digital processors, dithered to 16 bits if for the compact disc, then recorded into the DAW (dithering is explained in Chapter 15).

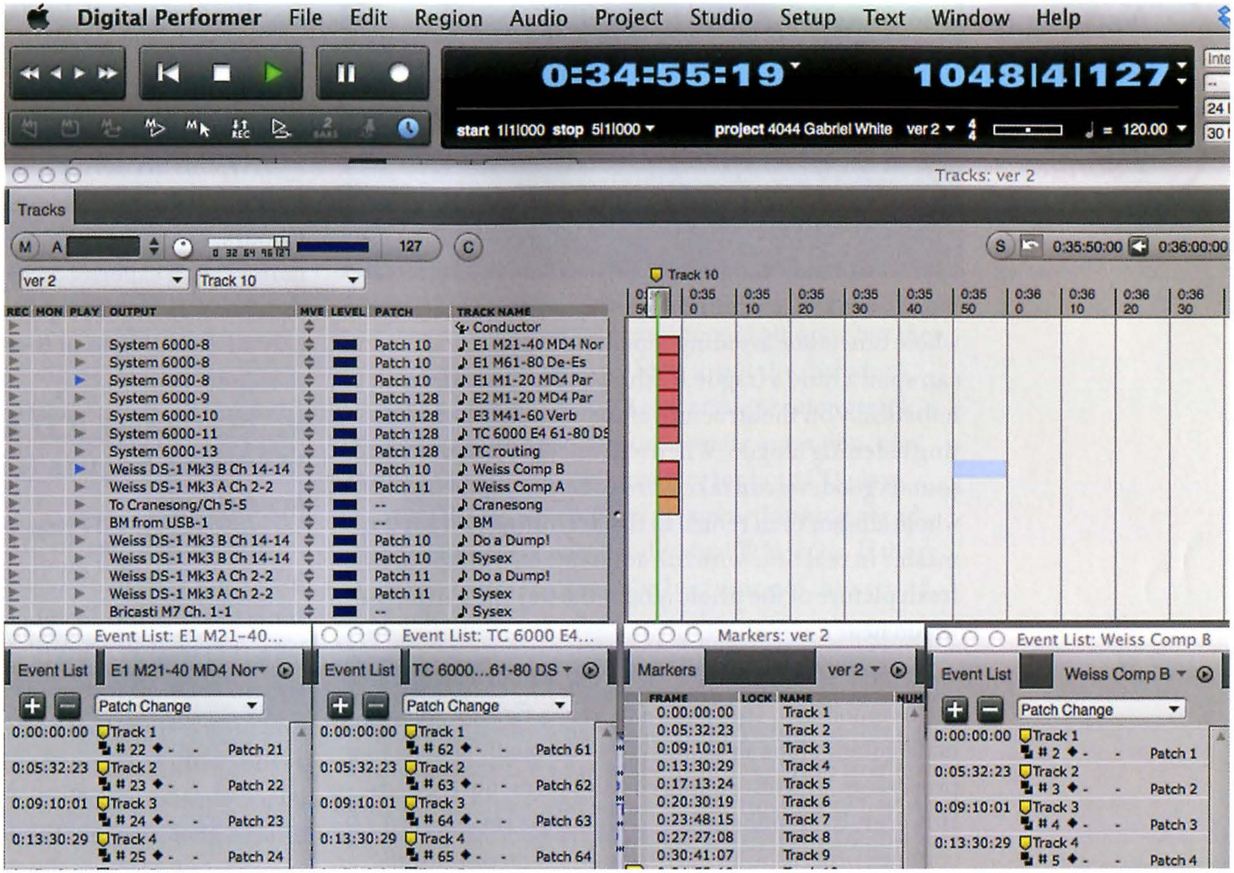
We master and capture one song, then move on to the next song, resetting processors until the best sound is achieved for that song, and so on as illustrated in **Var. 1** of the figure on page 48. In **Var. 1**, since all leveling, fading, and processing has already been completed, the DAW is used only for assembling and spacing, which is a very efficient approach. When we reach the end of the tune, if it requires a fadeout and we missed it, instead of reloading the entire song, we can back up before the fadeout, do a simple punch-in on the workstation, perform the fade, and then a matched edit. At this point, if the client orders any revisions, the engineer, to avoid going down a generation, must repatch the entire chain, reset the processors, make any processing changes, and re-record/replace the old destination file or a portion of the destination with a new one. For example, in the figure titled *Matched Edits* (page 49), the master is an assembly that begins with a piece of version 2, followed by version 1, a retouched section of version 1A, and finishes with version 1.5. *Retouched* refers to Cedar Retouch, a specialized noise reduction or restoration processor (See Chapter 8).

Often there is no real-time load-in, since sources usually arrive as computer files, and can be loaded at high speed directly into the workstation (**Var. 2**). After editing, assembly and cleanup, we proceed as in Var. 1, except using the workstation as the new "source" as well as destination.

In **Var. 3**, the mastering engineer waits until the final output to perform dithering, which gives some flexibility to perform fadeins and fadeouts on the final DAW file and perhaps some leveling, and simultaneously prepare 24-bit files for Mastered for iTunes (MFIT) and 16-bit files for CD (Chapters 15 and 16 cover dither and level issues in MFIT. For more details see my book *iTunes Music*). With the increasing number of high-sample-rate projects, an alternative is to play high sample rate material and capture it at the higher rate for higher resolution media or for LP. Yet another approach is to use **upsampling** followed by **downsampling**, because even if the source material is ready for CD at 44.1 kHz, digital audio processing at a higher rate sounds better (See Chapters 22 and 23). If the material is not already at the higher rate, the engineer may upsample the material in advance, process and capture at the higher rate and retain those higher sample rate files for other media, then downsample and dither for CD or digital downloads, which are currently at 44.1 kHz rate. Since a DAW can only work at a single sample rate at one time,[5] material that arrived at multiple sample rates (different songs at different rates) should be normalized to a single sample rate before the mastering can get started so all the clocks can stay stably locked to each other. Some engineers believe ADC sounds better than a sample rate converter, so they use an ADC at a higher rate as an effective upsampler.

Linear vs. Non-Linear Production

All of the above descriptions have one thing in common: they follow a *linear* approach to album mastering, capturing a song, then resetting the processors before moving on to the next tune. To ensure the context is good, it's best to go back to the end of a previously-captured song and play into the next as we capture it. Although engineers have been making excellent albums using this *linear* method for years, this method requires committing to the sound of the previous song before moving on to the next. Perhaps I want to achieve a slow crescendo from song to song with a climax in the middle leading to a dénouement. To achieve that, I like to have the ability to easily change any decision in a *non-linear* manner until I'm sure the whole album produces the effect I want. This avoids the drawbacks of the linear approach, which either forces us to recapture a song, or add another generation of processing. I like to review any song during the decision process and easily change it if it's going to help the impact of the whole album. **MIDI** is an aid to non-linear production. Most outboard digital audio processors and a small number of analog processors are remote-controllable via MIDI (**Var. 4**), which permits them to be automated and completely integrated with the workflow. The source DAW feeds timecode to a MIDI sequencer – in my case, a Macintosh computer running *Digital Performer* (pictured at right). The MIDI instructions are fed from Performer to the external rack processors. Performer can also be used to automate its own native plug-ins and act as an outboard processor, supplementing the CPU of the mastering DAW.



Digital Performer (MIDI sequencer) in action, controlling program changes in outboard gear

Many engineers already use automation in their work, since advanced workstations provide automated equalization, leveling, fades, dynamics, and plug-ins. If a later revision is requested, the mastering engineer can recall the previous EDL (edit decision list) and instantly make changes in any section of the album in the amounts or timing of the workstation's internal processing. *The MIDI technique extends this ability to the outboard equipment.* For me this is a revolution: finally, I can work on an album-in-the-making in a comfortable, fluid, non-linear manner. I work with a song until it is "cooked," save the parameters in the memories of the external processors, then move on to the next song, postponing the capture of the final file until the entire album has been programmed. I then return to near the end of the previous song and play the sequence, with the MIDI automation following along, nondestructively. This makes it easy to integrate two dissimilar songs, e.g., if one ends big and the other begins small. We can bring the album to a great climax, then recheck the first song in that context and instantly change its processing (if necessary) without having to reload or recapture. When we discover the introduction is too soft, but otherwise a song sounds fine, we can make that subtle adjustment before the capture

"Never count the seconds between songs."

time. Full automation facilitates creative special effects – for example, as I approach the climax on one tune, upon the entrance of a big vocal chorus, I can create MIDI-automated changes in an outboard processor that gradually increase the spaciousness and depth, producing a giant sound in the final chords.

The automated approach helps us keep things fresh. We can work on parts of a tune without having to play the whole tune, thus avoiding repetitive listening, which can spoil a tune's fragile freshness. We can concentrate holistically on the structure of the album without getting listening fatigue. When we conclude that the album sounds good, we can take a break before capturing the whole album, then return to the beginning and cut the master in real time with full automation. This gives us a fresh picture of the whole album the way the consumer would hear it.

If there is also analog processing, which usually cannot be automated (not shown in **Var. 4** figure), we may make minor stops along the way to manually change the processors, and afterwards edit the segments together. However, it is cumbersome to notate the settings of a complex processor for each tune and then capture all the tunes. If a complex non-automated processor must be changed from tune to tune, this makes non-linear production impractical.

The biggest advantage of full-automation is the ease of revision, especially if we have a critical clientele. Processing is applied in a non-destructive, non-cumulative manner so anything can be undone without going down another generation or forcing a reload. Another advantage of this method is that the raw sources can be immediately compared with the master and the difference demonstrated to the client. The disadvantage of this method is the learning curve required to run a MIDI sequencer and assemble a MIDI-controlled system.

How Long is a Mastering Session?

A typical mastering session for an album-length popular music project takes about a day, between 4 and 6 hours, rarely as much as 8. It may take an hour (rarely more) for an experienced engineer to patch and get the sound for the first song in an album. The second song may take a half hour or much less, generally reducing as the album goes on.

VI. PQ Coding (CD Albums) and Managing Segues (CDs and Downloads)

PQ Offsets

Most authorities on CD mastering recommend placing a track start mark (called **Index 1**) at least 5 SMPTE frames before the downbeat to accommodate slow-cuing CD players. This is approximately 12 CD frames, 160 ms (one CD frame = 1/75 second). The DAW can automatically apply these offsets, and show the PQ codes as they will appear on the disc. Sophisticated DAWs let us rehearse the effect of cuing with or without the offsets, critical when the cue has to be very tight.[6] For example, when the previous song is crossfading into the next, if we do not place the track mark extremely close to the downbeat of the next song the CD player may play a piece of the previous sound. I may accept as little as 2 (occasionally 1) SMPTE frames, which risks that a slow-cuing player will miss the downbeat. Pictured (page 53 at top) is an example of a live album with the track mark located nearly on top of the downbeat to avoid the spoken introduction. Some players clipped the downbeat, but on this CD it was less of a problem than hearing the previous sound.

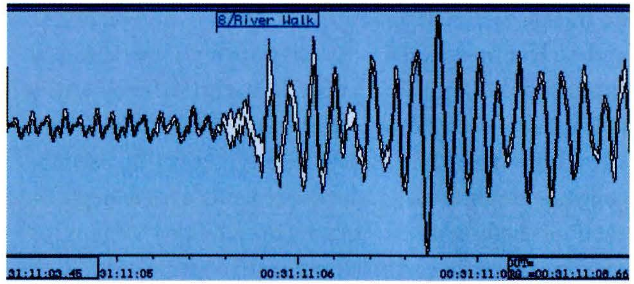
Spaces and PQ (Track) Coding

Index 0 is an optional mark between the tracks which defines the end of the previous track; the CD player's time display begins to count backward up to the Index 1. This is called the *pause time*, a misleading term, for there is no requirement for silence and in fact, index 0 can be 0 seconds. I recommend normalizing Index Zeros shorter than 2 seconds to 0 to keep the player's time display from glitching. This doesn't mean that musical spaces cannot be as short as you want them to be: it just means there will be no *official* pause between tracks. When Index 0 is 0 seconds, the player interprets Index 1 as the end mark of the previous track and the start of the next.

Hiding in the Gap. Segues and iTunes Singles.

When a cut from a concert album is played on the radio, it's often desirable to cue the tune on the downbeat: but the listener at home wants to hear the atmosphere between cuts, and maybe the artists' introductions. To accomplish this dual feat, the creative mastering engineer can place Index 0 and Index 1 times as in the figure (bottom right).

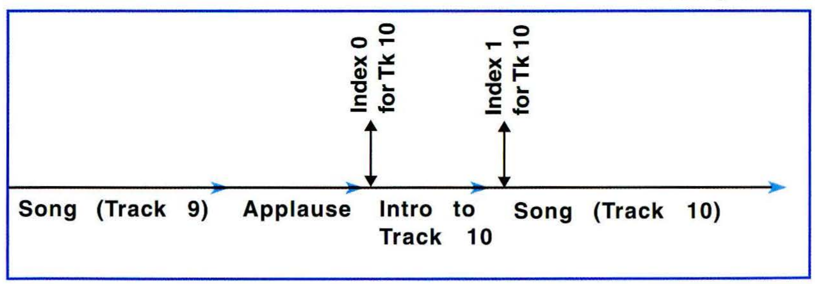
In this example, the song for track 9 ends with applause; the official end of song 9 is at Index 0. There is sound in the pause time between Index 0 and Index 1; this permits consumer choice or the CD player's random play function to ignore the boring or irrelevant parts. Similarly, the introductions, count offs, sticks, and so on, for songs on any kind of album can be placed in the gap so they will not be heard on the radio or in random play. The pause time does not count as part of the official length of either track (which keeps royalty costs down!). Unfortunately, this functionality of the CD standard is eroding, hindering artistry that we have enjoyed for over 30 years. iTunes and other computer players can play gapless (continuous) albums, but they do not read Index 0; so we have to put the introduction or the countoffs at the end of the previous track, producing some incongruous results in random play. Now we have to consider masters for both CD players and iTunes, which requires some thinking ahead. iTunes requires sending individual WAV files. If there is a segue between tunes, the beginning of the second tune will contain the fadeout of the previous one: so to generate singles, I have to capture the first song's ending in the clear and the second song's beginning in the clear to make two separate single WAV files for iTunes. The artist may not want to manage a separate single version of two songs, and so may reject the idea of having a segue. See my book *iTunes Music* for more examples.



Track mark placed very tight to the downbeat with no offset to avoid hearing talking which comes before the mark.

Redbook Limits

The Sony/Philips Redbook specifies all the parameters for an audio compact disc. A CD may have up to



Live album: Hide an intro in the Index 0 gap to permit radio play on the downbeat at Index 1, but let the consumer hear the full album or choose to skip the intro during random or manual play.

99 tracks, each having a minimum length of 4 seconds, and each of these tracks may have up to 99 indexes (aka subindices). Rarely do we code CDs with indexes, since many players do not support them and most people don't know how to use them. Classical engineers who used to code movements with indexes are using a track mark for each succeeding piece. There is no standard CD length; maximum length can be stretched to 80 minutes if the plant tightens the line spacing to the minimum Redbook tolerance – but not all players can play the outer tracks of these discs. Individual plants specify length limits on the order of 78:00 to 79:38 (check with the plant). Always record CDR masters in **Disc-At-Once** mode if the DAW gives more than one choice.

Hidden Tracks on CD and Digital Downloads

Find the hidden track is a little game that some producers play with the record-buyer. The mastering engineer can easily hide a track in a CD by inserting many short, blank 4-second "dummy tracks" at the end of the CD prior to the "hidden" one, which forces the listener to cue many times before he can reach it. Another method is to put several track marks within the "hidden" song, which causes ripping programs to break it up into pieces. Yet another way to hide a track is to have a track mark with no music for a minute or more. Most of these hidden track methods are gone with digital downloads (since there is no such thing as a track mark), though it's still possible to make a song file with dead space after a song, followed by another, unannounced song, which would raise alarms with every QC person between the producer and the consumer!

Some CD players have the ability to rewind in front of track one; this is called the pregap or first Index 0. One company claimed the rights to hidden tracks in the pregap, but it's not officially permitted by the Redbook standard, and many plants will not press CDs with a hidden track in the pregap. To the best of my knowledge, the master for a hidden track CD must be a CDR: it cannot be a DDP.

Metadata

During PQ coding, we also enter the metadata for a CD in the DAW: Title, artist, genre, ISRC code, EAN code, etc., which are used in CD Text (See Chapter 1). Metadata for iTunes and other computer players is entered manually (usually online) by the client.

Our *day in the life* is now complete! Take a deep breath and move on to *Equalization*, our first adventure in processing.

1
Segue (pronounced seg-way) – a crossfade or overlap of two elements. Webster's: proceed without interruption. Italian: seguire, to follow.

2
But ballads do not have to be raised as much as you think. Read about the acoustic advantage in Chapter 19.

3
Mix engineers follow similar practice, beginning the mix process at the loudest point of the song.

4
Well, this is true for CD mastering. But if you go back to the age of LP, the engineer was forced to cut an entire side in one continuous pass. If he stopped, he created a locked groove, which is yesterday's E32 error. A sophisticated LP cutting engineer would note settings for each tune and manually change her processors during the banding between each track. Equalizers were developed with A and B settings, allowing her to press one switch during the intertrack gap, and then leisurely preset the opposite equalizer for the next track. This is primitive, but roughly equivalent to the fully-automated process described here.

5
One workstation (Sequoia) can work at two rates simultaneously by opening two instances, each using different sound cards.

6
PQ lists only need to contain the offset (actual) times, the ones which appear on the CD master. The required offsets are: Initial pause mark 2 seconds in front of track 1's start mark, equivalent to placing track 1 at 2 seconds absolute time. Verify your DAW software performs one or the other automatically. Last end mark 2 seconds after music end, so the player can stop spinning without losing the last sound. Recommended: Each track mark a few frames before music start and each end mark a few frames after music end, to prevent player muting or missing part of the audio (though most players do not mute audio at all anymore).

CHAPTER 4

Equalization Techniques

I. Introduction

The First Principle of Mastering

The first principle of mastering is: *changing anything affects everything*. This means that mastering becomes the art of compromise: of knowing what is sonically possible, then making informed decisions about what is most important for the music. An EQ technique used in mastering can be crucially different from an apparently similar technique used in mixing. For example, when mastering, adjusting the low bass of a mix will affect the perception of the extreme highs. Similarly, if a snare drum sounds dull but the vocal sounds good, then the voice may suffer when you try to equalize for the snare.[1] These problems occur even between elements in the same frequency range. During mixing, bass-range instruments that exhibit problems in their harmonic range can be treated individually, but in mastering their harmonic range overlaps with the range of other instruments. A mix engineer can significantly boost a bass instrument somewhere between say, 700 and 2 kHz, but in mastering, even a small boost in this range can have detrimental effects. Or let's say we need to fix a bass drum problem: to minimize affecting the bass guitar it may be necessary to try careful, selective equalization to "get under the bass" at the fundamental of the drum, somewhere under 60 Hz. Sometimes we can't tell if a problem can be solved until we try, so it's best not to promise a client miracles – but then they're delighted when we deliver them!

> *"Practice is the best of all instructions."*
> — Chinese Fortune Cookie

II. What is a Good Tonal Balance?

Perhaps the major reason clients come to mastering houses is to verify and obtain a good tonal balance. But what, exactly, is a "good" tonal balance? The human ear responds positively to the tonality of a symphony orchestra that always exhibits a gradual high frequency rolloff on a spectrum analyzer – as will most good pop music masters. The amount of this rolloff varies considerably depending on the musical style, of course, so we use our ears, not the spectrum analyzer, as the basis for our EQ judgments.

Another key to effective mastering is that *everything starts with the midrange*. The fundamentals of the vocal, guitar, piano and other instruments must be correct, or nothing else can be made right. The message in the music – and more literally in radio, Internet and low-cost home systems – comes from the midrange. Listen to a great recording that's playing in the next room. The information still comes through despite the filtering of the doorway, carpets and obstacles. Then try filtering the recording below 200 Hz and above 5 kHz (like an old movie). A good recording will still translate.

The mastering engineer tries to make the sound pleasant, warm, and clear, if that is appropriate for the genre. While a master can deviate from this to provide a deliberately different color – for example, a brighter, thinner sound[2] – the mastering engineer limits excessive deviation from neutral, to ensure that the sound will translate well to the widest variety of playback systems and over the air.[3]

Specialized Music Genres

The symphonic tonal balance is generally a good guide for rock, pop, jazz, world, and folk music, especially in the mid to high frequencies. But some specialized music genres deliberately utilize very different frequency balances. We could think of Reggae as the symphony spectrum with a lot more bass instruments, whereas punk rock is often extremely aggressive, thin, loud, and bright. Punk voices can be thin and tinny over a fat musical background. If this straining of the natural fundamental-harmonic relationships is excessive and done for a whole record, most people find it fatiguing, but it can be interesting when it's part of the artistic variety of the record.

Be aware of the intentions of the mix

Equalization affects more than just tonality: it can also affect the internal balance of a mix. So a good mastering engineer must fully understand the intentions and needs of the production team. In fact, mastering equalization may help the producer's balance if his judgment has been inadvertently affected by a monitoring problem in the mix environment.

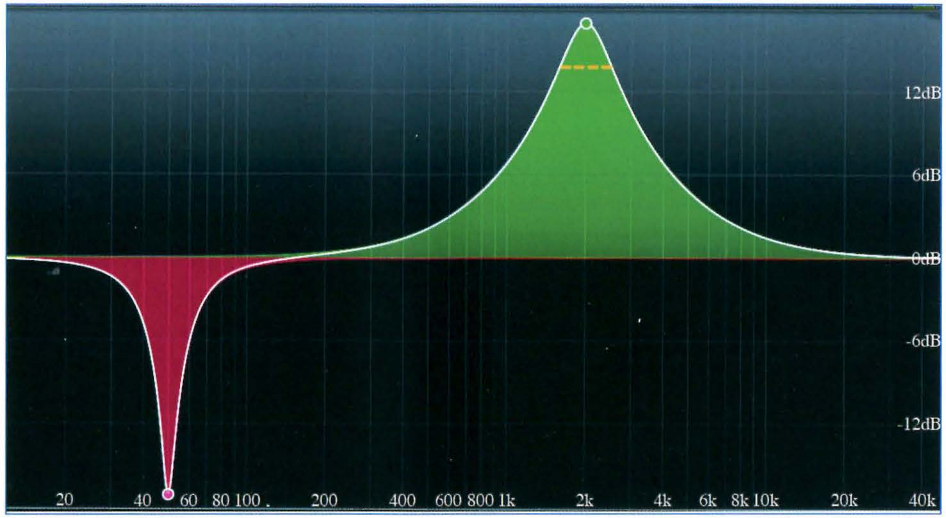
III. Equalization Techniques

Two Basic Types: Parametric and Shelving

There are two basic types of equalizers – **parametric** and **shelving** – named after the shape of their characteristic curves. Parametric EQ, invented by George Massenburg in 1967,[4] is the most flexible curve, providing three controls: center frequency, bandwidth, and level of boost or cut. Mix engineers like to use parametrics on individual instruments, boosting to bring out their clarity or salient characteristic, selectively dipping to eliminate problems, or, by virtue of the dip, to exaggerate the other ranges. The parametric curve, also known as **peaking** or **bell curve**, is also the most popular EQ shape used in mastering, because it can be used "surgically" to remove certain defects, such as overly-resonant bass instruments, or enhance narrow ranges of frequencies. By comparison, shelving equalizers are more popular in mastering than in mixing, since they provide boosts or cuts to the entire spectrum below or above a selected frequency, and can alter the tonality of the entire mix. With a good monitoring setup, equalization changes of less than 1/2 dB are audible, and in fact, a shelving high frequency boost of only 0.1 dB starting at 2 kHz is clearly audible because it changes the whole tonal curve of the material.

Parametric: Q and Bandwidth

The parameter **Q** is defined mathematically as the result of dividing the center frequency by the bandwidth in Hertz at the 3 dB down (up) points measured from the peak (dip) of the curve. A small value of Q means a large value for bandwidth, and vice versa. For example, a Q of 0.6 is equivalent to a wide bandwidth, 2 octaves, but a Q of 4 means the bandwidth is narrow, 0.3 octaves. There is a conversion calculator in the links. The figure (above right) shows two parametric bands with extreme levels for purposes of illustration: on the left, a 17 dB cut at 50 Hz with a very narrow Q of 4, which is 0.36 octaves or a bandwidth of 12.5 Hz; on the right, a 17 dB boost centered at 2 kHz, with a fairly wide (gentle) Q of 0.86, which is 1.6 octaves. The bandwidth is 2325 Hz, represented by the dashed yellow line.[5]

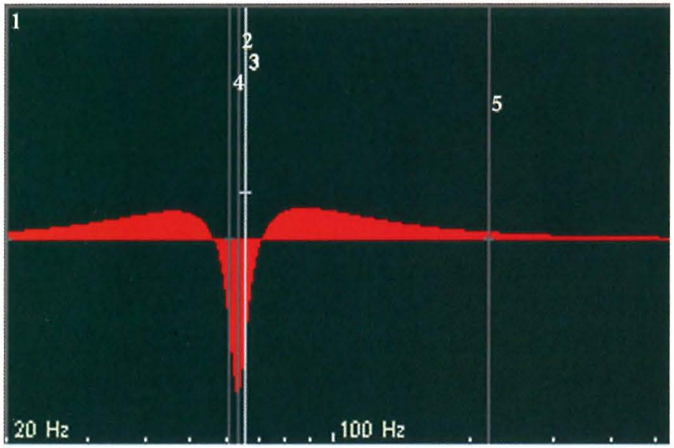


Parametric equalizer with a cut of -17 dB at 50 Hz with a very narrow bandwidth of 0.36 octaves (Q = 4), and +17 dB boost centered at 2 kHz with a fairly wide bandwidth of 1.60 oct (Q = 0.86), indicated by the dashed yellow line at the 3 dB down points.

Gentle equalizer slopes almost always sound more natural and less harsh than sharp ones, so Q's of 0.6 and 0.7 are therefore very popular in bell shapes and gentle slopes in shelving shapes. Higher (sharper) Q's (greater than 2) are used surgically, to deal with narrow-band resonances or discrete-frequency noises, though we must listen for artifacts of high Q, such as **ringing**. It is possible to work on just one note with a sufficiently narrow-band equalizer, or we may overturn the first principle of mastering by using a higher Q to try to isolate and emphasize a single instrument. For example, a poorly-mixed program may have a weak bass instrument. Boosting the bass at around 80 Hz may help the bass, but it might also muddy the vocal. So we try narrowing the bandwidth of the bass boost. This is rarely totally effective, so if the bass boost is not good for the vocal, it's probably not good for the song. But if the vocal is made only slightly bassier, we can try a slight compensatory boost around 5 kHz, as long as that doesn't interact poorly with yet another instrument!

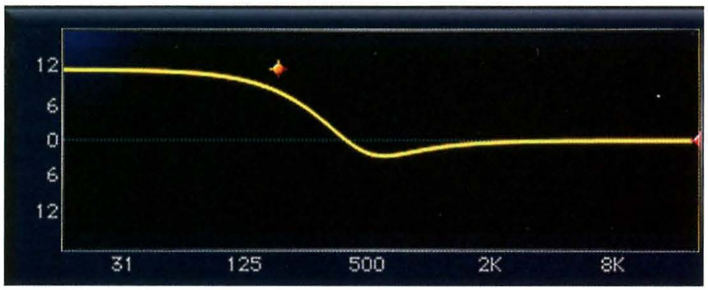
Finding the right EQ frequency for dipping resonant notes

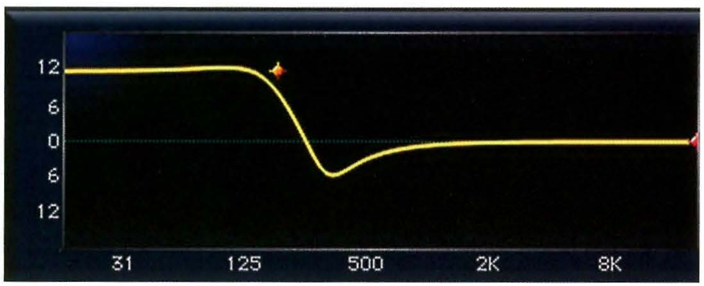
There are two techniques for finding a problem frequency that is resonating and must be dipped. The *classic approach* is to *focus the equalizer directly*: starting with a large boost and fairly wide (low value) Q, sweep through the frequencies until the resonance is most exaggerated. Then narrow the Q to be surgical, and finally, dip the EQ the amount desired. This technique works



"One-note bass" resonance fixed by a combination of a shelving boost (which was useful to help the rest of the notes that were weak) and a narrow band dip at the resonance frequency.

well with analog equalizers, but some digital equalizers present ergonomic obstacles: the inefficient mouse, and latency. It's very difficult to sweep in the bass region because the distance between F# and G is only 3 Hz, while in the midrange, the distance is 22 Hz. We've all heard the phrase "one-note bass", and there's a reason why this problem occurs: many rooms have standing wave problems in the bass that give the mix engineer the wrong impression that a note is too weak. So he boosts it unnecessarily. The cure for one-note bass on the mastering side is quite delicate: we have to construct a bell filter that's only a few Hz wide. Narrow filters can ring, so be exceptionally careful: if the bass player doesn't play the adjacent notes, don't make the filter any narrower than necessary, and only dip as much as is necessary. In a DAW it is also possible to automate the EQ so it occurs only when the offending note is playing. This gives rise to a *second technique* specifically for bass frequency surgery: keep a keyboard handy to determine the key of the song, and use your sense of relative pitch to determine the problem note. Then translate that note to a particular frequency with the Carnegie Chart (attached to the front of this book) and dip that frequency. Pictured (top left) is an example of a bass EQ found in just seconds using this method. It combines a shelving boost with a single dip at the problem frequency. Engineers with a *Crane Song Ibis* EQ or with the *DMG Equilibrium* can skip the chart, because the Ibis is marked directly in musical notes, and the Equilibrium includes a synchronized graphic of a keyboard (pictured on page 60). Why didn't designers think of that sooner? With good ergonomics we can work faster and finish earlier.

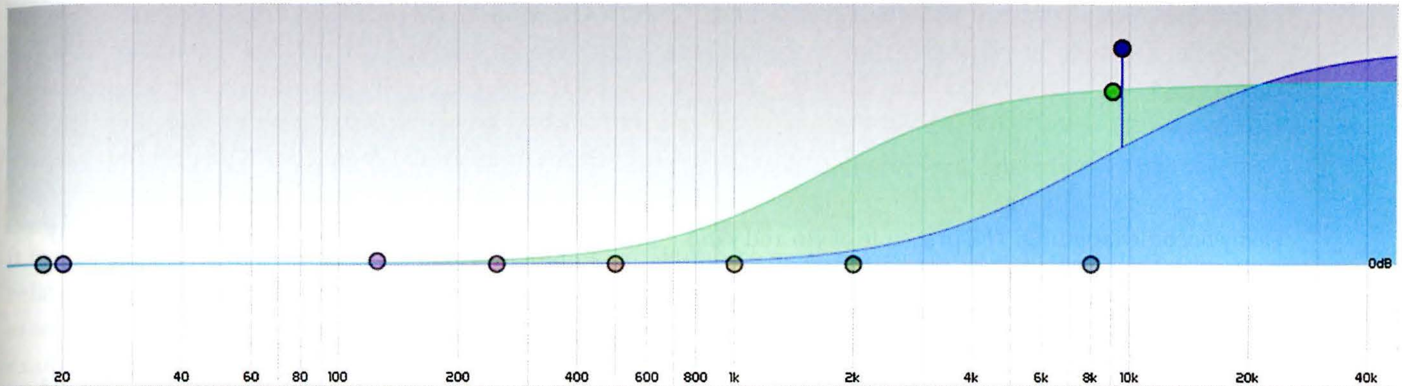




Top: Gerzon resonant shelf with a low Q. The dip just past the shelving boost frequency is characteristic of the Gerzon resonant shelf. Bottom: The same with a high Q.

Shelving Equalizers

As mentioned, a shelving equalizer affects the level of the entire low frequency or high frequency range below or above a specified frequency. Some have Q controls, defined as the slope of the shelf at its 3 dB up or down point. One interesting variant on the standard shelf shape can be found in the DMG Equality, Manley Massive Passive, Waves Renaissance EQ, and Weiss EQ-1. This shape, called a *resonant shelf* (two pictures lower left), was proposed by psychoacoustician Michael Gerzon. I like to think of it as a combination of a shelving boost and a parametric dip (or vice versa). In the top image, a low Q (0.71) bass shelf below 178 Hz is mollified by a gentle parametric dip above 178 Hz; all of which is controlled by a single band of the equalizer. This type of curve can help keep a vocal from sounding thick while implementing a bass boost. The bottom image shows the same boost with a high Q of 1.41.



Gentle Baxandall curve (blue/violet) vs. "standard" shelf (green). To approximate a Baxandall shape try a shelf with 3 to 6 dB per octave slope.

EQ Yin and Yang

Remember the yin and the yang: contrasting ranges have an interactive effect. For example:

- Adding low frequencies makes the sound seem duller, and reducing them makes it seem brighter.
- Adding extreme highs between 15-20 kHz makes the sound seem thinner in the bass/lower midrange, and vice versa.
- A slight dip in the lower midrange (around 250 Hz) reduces warmth, and has a similar effect to boosting in the presence range (around 5 kHz).
- A harsh-sounding trumpet section can be improved by dipping around 6-8 kHz, and/or boosting ~250 Hz.
- A thick vocal can be helped either by reducing the lower midrange, or adding presence or both

Yin and yang considerations allow us to work in either or both contrasting ranges, whichever is most effective. When the overall level is too high, pick the range you need to reduce. When an instrument(s) exhibits upper midrange harshness, pick the frequency range that will have the least effect on other instruments playing at the same time.

Using Baxandall for air

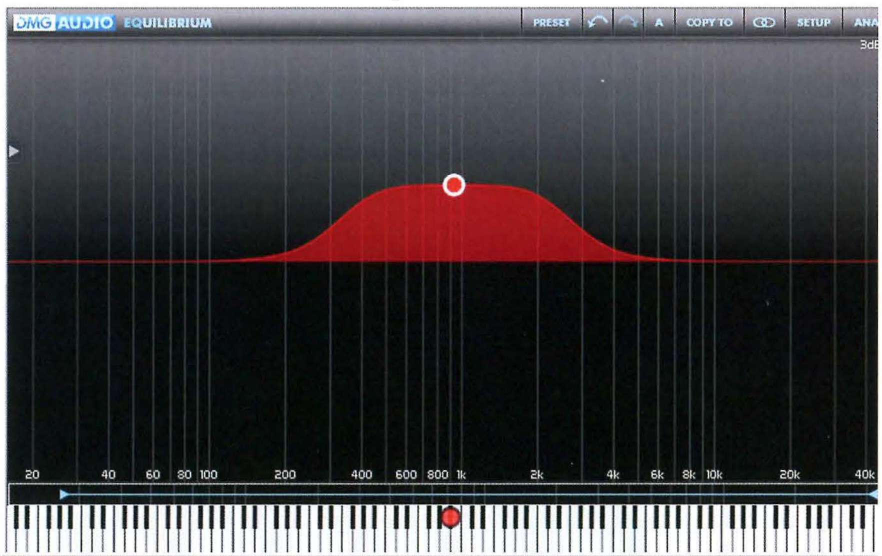
In Chapter 2, we described the *air band* as the range of frequencies between about 15-20 kHz — the highest frequencies we can hear. An accurate monitoring system will indicate whether these frequencies need help. An *air boost* is contraindicated if it makes the sound harsh, or unintentionally brings instruments like the cymbals forward (closer or louder). A third and important shape that's extremely useful in mastering is the **Baxandall** curve (blue/violet curve pictured above), named after Peter Baxandall. Hi-Fi tone controls are usually modeled around this curve. Like shelving equalizers, a Baxandall curve is applied to low or high frequency boosts or cuts. However, instead of reaching a plateau (shelf), the Baxandall continues to rise (or dip, if cutting instead of boosting). This gentle shape often meets the ear's desires better than a "standard" shelf, especially for whole mixes, which is why Baxandall is very popular in mastering.

As a comparison, pictured above is an API-style shelving boost (green); true to its name, from about 6k on up it is a plateau, while the Baxandall curve (blue) continues to rise, even above 20 kHz (violet portion). You can simulate a Baxandall shape using an EQ with a variable slope or Q. We can also simulate a Baxandall high frequency boost by placing a parametric equalizer (Q= ~1) at the high-frequency limit (~20 kHz), ignoring the right hand portion of the bell curve.

Be careful when making high frequency boosts (adding *sparklies*). They are initially seductive, but can easily become fatiguing. The principle of yin and yang reminds us that the ear interprets a high frequency boost as a thinning of the lower midrange or bottom. In addition, when the highs come up, the cymbals, triangle and tambourine become louder, which changes the balance of rhythm to melody, for better or worse.

It's All about the Curve

Here are two other curves which we don't ordinarily see, but which are available in the DMG Equilibrium. Below is a Butterworth-shaped bell boost, with a large plateau, useful when a broad section of the midrange needs a sharp boost. I confess I haven't tried this one.

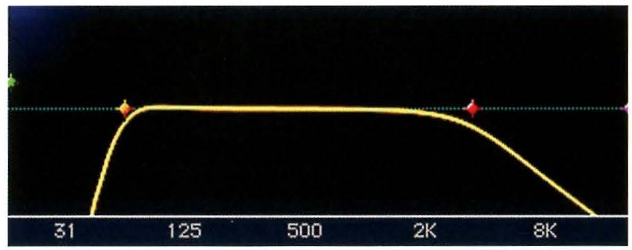


Butterworth boost

The curve on the top of page 61, in the DMG Equilibrium, is called a *tilt curve*. Tilt curves are very useful for mastering. If a recording is a bit bright overall or a bit bassy, tilt will fix it with a single filter. The pictured example shows an 0.25 dB bass tilt centered at 1 kHz. Since tilt affects the yin and the yang and extends to both frequency extremes, even 0.1 dB tilt is clearly audible. Tilt is equivalent to two first-order shelves located "back-to-back".

High-Pass and Low-Pass Filters

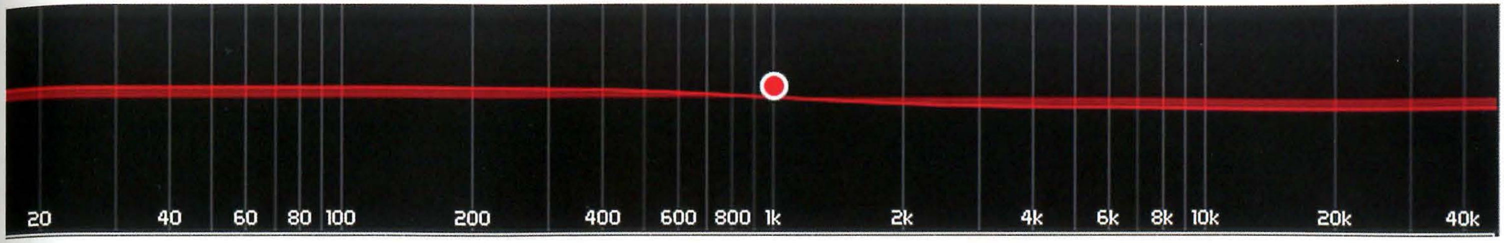
On the left of the figure below is a sharp high-pass (low cut) filter at 61 Hz, and on the right, a gentle low-pass (high cut) filter at 3364 Hz. The frequencies are defined as the points where the filter is 3 dB down. Al-



though pass filters can be used to solve noise problems in mastering, they can also introduce problems of their own, because they affect everything above or below a certain frequency. High-pass filters can reduce rumble, thumps, p-pops and similar noises. Low-pass filters are sometimes used to reduce hiss, but since the ear is most sensitive to hiss in the 3 kHz range, a parametric dip around that frequency is more effective than the radical low-pass filter. For hiss removal, we usually prefer specialized noise-reduction solutions over static filters (See Chapter 8).

One channel or both (all)?

In most cases, applying the same EQ adjustment to both (all) channels is the best way to proceed, because



0.25 dB bass tilt. It looks subtle, but it's distinctly audible because it affects the entire spectral character.

it maintains the stereo (surround) balance and the relative phase between channels. But sometimes it is essential to be able to alter only one channel's EQ. For example, with a too-bright high hat on the right side, a good vocal in the middle and proper crash cymbal on the left, the best solution is to work on the right channel's high frequencies. In Chapter 9 we will discuss MS manipulation of equalization.

Start subtly first

Sometimes important instruments need help, though ideally, they should have been fixed in the mix. The best repair approach is to start subtly and advance to severity only if subtlety doesn't work. For example, if the piano solo is weak, try to make the changes surgically, and make them:

- only during the solo
- only on the channel where the piano is primarily located, if that sounds less obtrusive
- only in the frequency range(s) that help: fundamental, harmonic, or both
- as a last resort, by raising the entire level, because it would affect the entire mix, though the ear focuses on the primary instrument

The Limitations and the Potential of the Recording

If you wait until the mastering stage to fix certain problems, this invites a compromise, because there is only so much that can be done in mastering. But sometimes mix engineers try to fix things that didn't need repair, or overprocess a recording, only making it sound worse. They do this because: the tool is available and it's tempting to use; their monitoring is misleading; or because of lack of experience (the same thing can happen to an inexperienced mastering engineer). A plethora of plug-ins does not turn someone into an audio engineer. This is where it pays for the mix engineer to consult with an experienced mastering engineer before the mix is done. There is little we can do to fix a recording where one instrument or voice requires one type of equalization, and the rest require others. In many cases I recommend a remix. However, if a remix is not possible, then we might be able to use the specialized techniques to be discussed in Chapter 9.

As we discussed in Chapter 2, comb filtering is a complex problem that is not easily cured with an equalizer. Besides, in mastering, EQ affects the entire mix, not just the offending instrument or voice. It's best to first discuss the problem with the mix engineer to see if he can address the offending track and remix. If that is not possible, then possibly ask for a stem, or as a last resort try an overall mastering EQ — for example, a lower midrange EQ boost to help a vocal that sounds thin due to comb filtering, even if it only touches one band of the comb filter.

> *"The worst part of my job is trying to repair what others have fixed."*
> — George Guerin

The Story of One Hip Hop Album

The equalization I chose for one hip hop album has a deep shelving filter below about 100 Hz (or possibly below 70 Hz) that we can boost or cut as required to help the deep singing low bass note or *bass drop*. If the drop in the original mix needs help, using this EQ can raise or lower it a bit, keeping in mind that an equalizer works on frequency ranges, not on tracks, per se. The mix I received was quite conservative, and the deep bass drop note benefitted from being raised. In this case, since I chose to use a shelving boost, the lowest, rumbly part of the note (below about 30 Hz) became a bit too strong, so I applied a subtle high-pass filter as well. To zero in on the bass drum thump, I found a bell curve around 70 Hz (where the prime resonance or "essence" of this bass drum lay — though 60 Hz is a more common bass drum EQ frequency) with a fairly narrow Q (about 2). Usually I do not sweep, but sweeping helped confirm the best boost frequency for this bass drum to be 70 Hz.

The male vocal sound needed a bit of body, and since he was the prominent instrument in the 250 to 400 Hz region, a small bell curve boost around 250 Hz with a Q of 0.7 served to fatten up the vocal. The presence of the vocal, the synth, and the percussive

⟶ page 63

A perfect mix needs no mastering processing at all! Because of this, don't automatically begin equalizing, but listen and evaluate first. Many recordings that sound great leave the mastering studio with no equalization or processing.

Loud and Soft Passages

I almost always begin mastering with the loudest part of a song. Why? Because loud passages accentuate those peaks that the ear is most sensitive to. Equalization choices that are pleasing during a mezzo-forte passage may sound harsh during a loud one. We'll talk about the psychoacoustics of loudness in Section IV below.

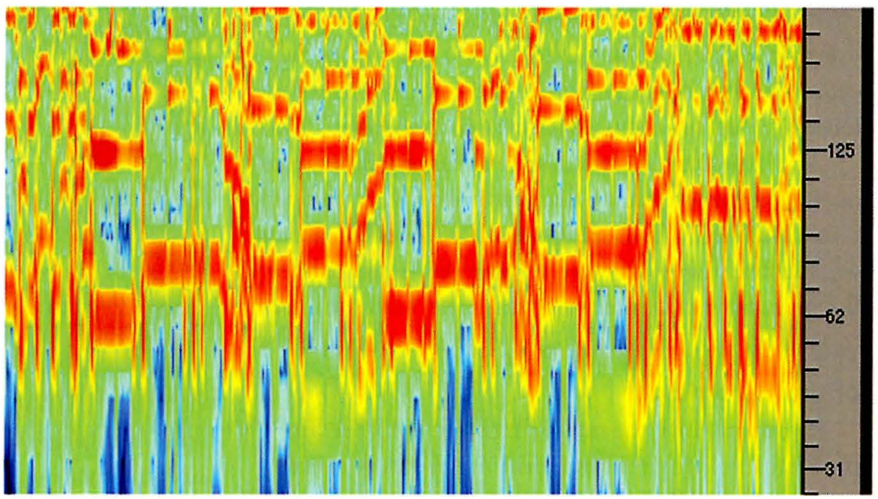
Fundamental or Harmonic?

The extreme treble range mostly contains instrumental harmonics. Since the fundamental of a crash cymbal can be lower than 1.5 kHz, boosting the harmonics too much makes a cymbal sound tinny or thin. When equalizing or processing bass frequencies, it is easy to confuse the fundamental with the second harmonic. This detail shot of a SpectraFoo™ **Spectragram** illustrates the importance of the harmonics of a bass instrument. High amplitudes are indicated in red, and descending levels in orange, yellow, green, then blue. Time passes from left to right like in a concert score.

Notice the parallel run of the bass instrument's fundamental from 62-125 Hz and its second and third harmonics from 125-250 and up. Should we equalize the bass instrument's fundamental or the harmonic? It's easy to be fooled by the octave relationship; the answer has to be determined by ear — sometimes one, the other, or both. Sometimes I EQ below the fundamental, even as low as 40 Hz, which can make the instrument sound fatter and richer even though there is no subharmonic. Perhaps this works because the equalizer I used has some harmonic distortion.

Bass, The Final Frontier[6]

Since the ear is significantly less sensitive to bass energy compared with the rest of the sound spectrum, bass information requires a lot more power for equal sonic impact: around 6 to 10 dB more below about 50 Hz, and about 3-5 dB more between 50 and 100 Hz.[7] This explains why bass instruments often have to be compressed to sound even. It also means that a low frequency boost introduces so much energy that it can reduce the highest clean program level we can give to a song (in cases where the client is demanding a "loud" master). Fortunately, the ear's tendency to supply missing fundamentals (see Chapter 2) works in our favor, allowing us to save "energy" by cutting with a fairly



SpectraFoo™ spectragram of the bass frequencies of several measures from a rock piece. Read it like an orchestra score, time runs from left to right. Red represents the highest levels. The bass runs in the 62-125 Hz fundamental range are paralleled by second and third harmonics.

sharp high-pass filter, but ideally only if it does not hurt the quality of the bass drum or the low notes of the bass. The high-pass filter must be extremely transparent and have low distortion. Sometimes a gentle filter is a better choice than a steep one, as when dealing with a boomy bass drum or bass. But subsonic rumble or thumps benefit from a steep filter to minimize the effect on the instruments.[8] Don't automatically high-pass a recording: always listen first. Sometimes a bass drum sounds better with full energy, sometimes not. Always think musically, e.g., listen in particular to the musical interaction between the bass drum and the bass.

The **Equal Loudness Contours** can seriously affect the level of bass and bass drum in a mix, because the ear is much less sensitive to low frequencies, and at louder levels it is a bit more sensitive. Thus we have to be careful that perceived bass level can come up during mastering simply because we've made the sound louder. When I listen to a mix that I intend to make louder in mastering, I mentally prepare myself to rebalance the bass, at least a little bit. Mixing engineers should try to audition their material at the intended listening loudness to make sure the bass response is in the ballpark. Also, mixing and mastering engineers should collaborate to ensure the bass balance is going to be correct when finally mastered. It's better to anticipate this issue during the mixing, because rolling off the bass during mastering is more of a compromise when we can't control the individual instruments' levels. For example, sometimes the bass instrument sounds correct but the bass drum does not.

This issue is aggravated when mix engineers use small loudspeakers, running the risk of producing an inferior product. Accurate subwoofers let us hear low frequency leakage problems that tend to muddy the mix — for example, bass drum leakage in vocal and piano mikes. It's much better to apply selective high-pass filtering during the mixing process, because mastering filters will affect all the instruments in a frequency range. For example, mix engineers may get away with a steep 80 Hz filter on an isolated vocal, but that's too high a frequency for mastering a whole mix.

IV. Other refinements

The Psychoacoustics of Loudness

Psychoacoustician Jim Johnston helps explain the physiological science behind loudness and in this particular case, equalization. JJ explains:

> Keep in mind that loudness is a perceptual term, as opposed to intensity or level, which is a measurable quantity. "Partial loudness" refers to the part of loudness sensation that is due to any given frequency range, typically given across a bandwidth of 1 critical band, with spacing of the centers substantially closer than 1 critical band. Loudness is the sum of the partial loudnesses. What we reduce in our central nervous system to "what we heard" is the result of partial loudnesses over time. EQ peaks (boosts) affect the partial loudness greater than EQ dips, because of how the central nervous system works.

This nonlinear behavior of the central nervous system explains why louder sounds better and why our ears are more attracted by EQ boosts than dips (at least in an instant comparison).

The Effect of Psychoacoustics on EQ choices. Let's look at equalization from this new point of view. Equalizers affect partial loudness in the band that is being manipulated. Is it possible that much of our preference for one EQ setting over another is simply that it

instruments are all married together in the 5 to 8 kHz region, so if some of them need help with a boost, and others are too prominent, then a single EQ would be ambiguous. Let's say this vocalist needs some presence (clarity), as does (fortunately) the synthesizer. But the percussive claps and noises are too "sharp" sounding or hard sounding. We won't get anywhere by trying to both boost *and* cut in the same range. By using a single band compressor (see Chapter 6) in the same frequency range I was able to soften the percussive bite yet still help the presence of the vocal and synth with the EQ.

The last band I engaged is a high frequency shelf (10 kHz and up), which helped the air frequencies of the cymbals and some of the vocal presence. If the vocal gets too much presence with the extra HF boost for the cymbals, I would try just shelving up the highs in the S channel (see Chapter 9), since the vocal is usually centered and the cymbals are often more at the sides of the stereo image. Keep in mind the art of compromise: if any of these choices helps one instrument while hurting another, pick the most important instrument: see if the ear focuses on that and ignores the negative effects on the other instruments caused by that EQ choice. Or, don't use so much EQ — a very wise choice!

sounds louder to the ear? Does the loudness at which a clip is played affect our judgment of an equalizer? The answer is, yes, the loudness matters *very* much! For example, let's say I raise the 5 kHz band by a small amount and it seems to create a better impression of a recording. If I then carefully lower the gain until the perceived loudness is the same as the flat signal – the impression changes dramatically. Try it yourself: it's quite surprising. The 5 kHz band, which was originally making the recording seem a bit louder, is now reduced in absolute loudness, as are the other bands, and when instantly comparing the two EQs, the new impression is that the sound is warmer: there is less presence, completely changing our impression of the equalizer. As mastering engineers we have to ask ourselves if it is really the tonality or simply the instantaneous loudness difference that we prefer. Thus, all equalizers should have gain controls to let us make the loudness in bypass match the loudness with the equalizer engaged. This will guard against unnecessary equalization during instantaneous comparisons. Even without gain compensation, it is good equalization practice during mastering to avoid switching the equalizer rapidly in and out, while making instant judgments. Instead, play the whole passage once with the equalizer inserted and once bypassed. This will help us to make a more objective, long-term judgment. It turns out that equalization is far less important than loudness management itself.

For example, let's consider a subtractive EQ, which uses dips instead of boosts, and see why some engineers react to a subtractive EQ by saying that the sound seems to "lose quality," but cannot explain why. The answer lies in the psychoacoustics of loudness.

This EQ curve (below) shows a dip in the lower midrange.

If we consider the grey line in the next picture as 0 dB, the unity gain line, then an EQ dip will lower the partial loudness of the sound.

Thus we perceive the sound as less "vivid", though it is really just a little less loud (softer). When we apply some gain to our curve, using the identical equalization and placing the bottom of the EQ dip at unity gain, then every other frequency has been raised proportionally.

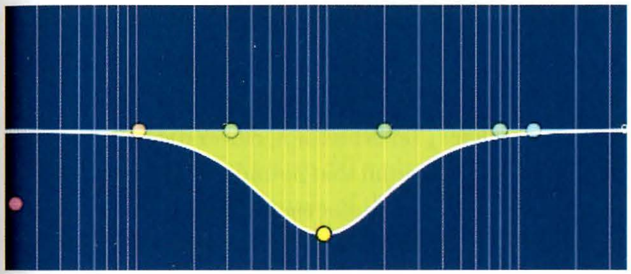
Our brains will now hear a bass and treble boost instead of a midrange dip – and an overall loudness boost! Our perception of this EQ is very dependent on the loudness context: that is, how quickly it is auditioned after the flat version, how loudly we play the flat version, and how loudly we play the equalized version!

When we raise the gain just enough to match the loudness with the flat sound, the loudness disadvantage of the subtractive equalization goes away and we can judge the EQ setting on an equal footing.

Be aware that it is extremely difficult to exactly match the loudness of different equalizations – so the grand illusion is always going to be part of the job of adjusting an equalizer. I'm not suggesting you obsess over this discovery: just be aware of the loudness factors that influence your sonic decisions. Keeping these principles in mind, the underrated subtractive EQ will become as important a part of our creative palette as additive EQ.

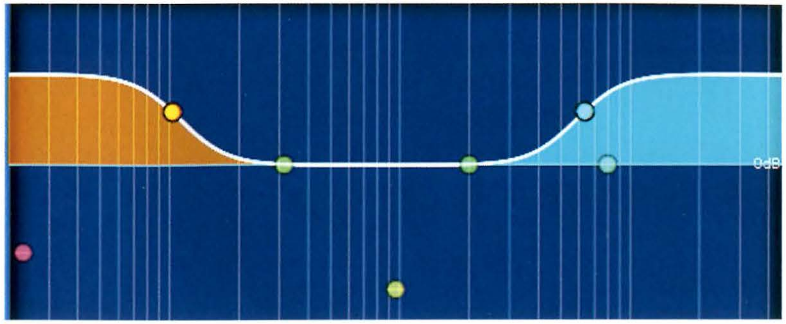
Shaping the Shapes

The shape of the curve is very important to perception and the practice of equalization. Pictured (below) is a midrange dip, using a single bell curve. Notice that the high and low frequencies are flat. If the center of the bell were flat instead of curved, it might be perceived as a bass and treble boost instead of a midrange dip.



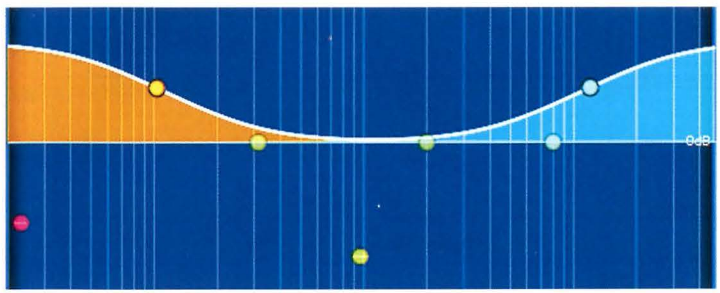
Simple midrange dip. Notice that the high and low frequencies are flat.

Next is a curve (top right) created by boosting a "classic" low and high shelf. Notice how much the middle of this curve resembles the above midrange dip, except that the middle is flat. We do hear a brightening and "bassy" effect, but primarily we notice a "presence boost" [9] because the rising (diagonal) portion of the shape tickles the ear. This is the shape to use if you're looking for upper mid-frequency presence more than sparklies at the high end, or mid-bass tone more than deep bass punch at the low end. We notice the sloped portion of the curve more than the flat portion at the frequency extremes, since the flat portion becomes a sort of reference for the ears, even though both extremes are boosted entirely above the midrange.



"Classic" low shelf and high shelf boost. Notice the resemblance of the flat-topped frequency extremes to those in the midrange dip.

Here is a "smiley curve" (below) created by using two shelving boosts with adjustable shape, set to a gentle slope. The smiley curve extends the sonic effect all the way to the frequency extremes, because the curves are never flat. It can simultaneously produce punchy deep bass and sparkly treble. No wonder it's so popular! Admittedly, one of the reasons this shape is popular is because it sounds louder on an A/B comparison with the original (unless the gain is compensated). Contrary to the classic shelf, the high boost of a smiley curve can sound very smooth in the upper midrange, almost invisible, because it works gently and continuously.

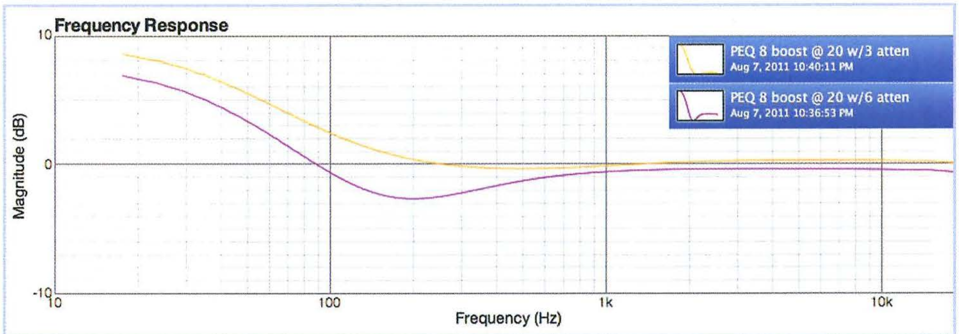


A "smiley curve" formed by two shelves with adjustable shape. It resembles a low and high Baxandall boost.

"Mastering is the art of compromise."

The Pultec Curve

Pultec is a brand name of the company *Pulse Techniques*, formed in the early 1950s and no longer in existence. The much-revered Pultec equalizers have been revived in a few pieces of hardware, and in some plug-ins that emulate the distortion characteristics of its analog electronics and reproduce its unique curves. The frequency responses (pictured below) were obtained by combining a large Pultec low-frequency boost with a small low-frequency cut, both at 20 Hz. In the red curve, it appears as if an additional lower-midrange dip has been added, as if by magic. But the real explanation is that the cut frequency is purposely offset from the boost frequency by 100 Hz, so a bass boost and cut of exactly the same amount would have flat response in the bottom, but would create a dip in the lower midrange. It is a versatile trick, giving the effect of three bands with only two controls. A Pultec can be used to create many other nice effects, which you can learn about by reading the Bettermaker User Manual (which I helped write).



Pultec response curve obtained by simultaneously applying bass boost and bass cut. The violet curve has more bass cut than the yellow.

Analog Equalizers

Analog equalizers are still widely used by mastering engineers, in part because of the ergonomics, since knobs are much easier to use than computer mice. That said, outboard digital and digitally-controlled equalizers like the Weiss and the Bettermaker combine analog-style ergonomics with digital control flexibility – such as A/B switching, memories, and so on. Analog equalizers are also popular because they can easily produce a warm sound, compared with the early digital equalizers, which sounded somewhat "sterile." But digital designers have learned a lot and this is no longer so true. (We discuss this in Chapter 22.) In addition, the curves of analog equalizers are not constrained by the **Nyquist** frequency, so they remain more natural at the high frequencies. Yet even then, clever digital designers have found ways around that boundary. The advantage of *digitally-controlled analog equalizers* is that they combine the power of digital with the sound of analog. There are only two extant models to my knowledge: The *SPL*, which has motorized faders, and the *Bettermaker*, which uses active-controlled circuits and has excellent sound and ergonomics (pictured page 67, top). I won't get into the sonic differences between these two models. Let me just say that you should not leap to conclusions, and carefully audition each one.

Linear phase Equalizers: The Theory

All current analog equalizer designs, and nearly all current digital equalizers, produce phase shift when boosted or cut: that is, signal delay varies with frequency, and the length of the delay changes with the amount of boost or cut. The higher the Q, the more the phase shift. This kind of filter will always alter the musical timing and wave shape, also known as *phase distortion*.

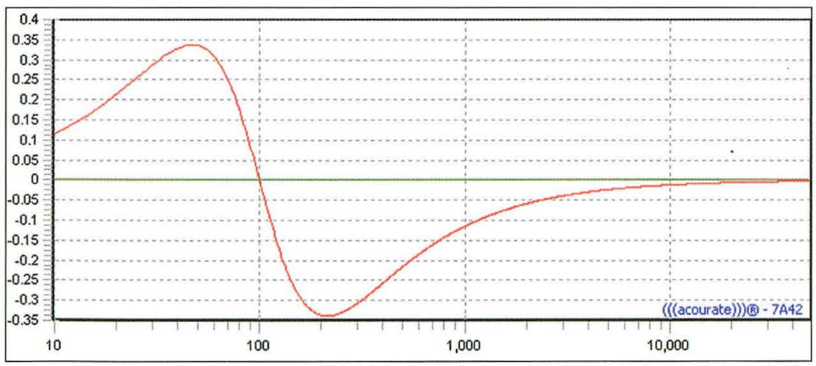
Bettermaker *EQ 230P* *digitally-controlled analog equalizer. Only two models of digitally-controlled EQs exist to my knowledge. This model has 399 memories, a remote-control program and contents can be stored via sysex.*

However, we've grown quite used to the sound and effect of phase distortion. Let me elaborate.

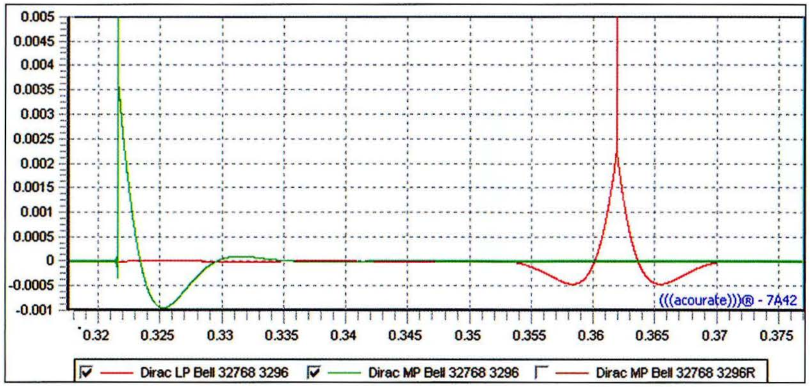
In the figure (bottom right) we graph the phase response of two equalizers, each with a bell curve boosted at 100 Hz. Vertical scale is phase in radians, horizontal scale is frequency. In red is a minimum phase (MP) EQ, and in green a linear phase (LP). Notice that the linear phase EQ's phase response is a straight line, and the phase of the MP takes a nose-dive at the center frequency of the boost. What does this phase shift actually mean?

Jim Johnston outlines the fundamental technical differences:

> Whenever you have to equalize, you will alter the signal in both the time and frequency domains (as mathematics requires); there will always be a *time artifact*. In the *analog* style equalizer, which is usually mathematically termed *minimum phase*, the alteration will be primarily to spread the signal *downstream*, i.e., it does not lead the original signal by much, if any. A *downstream modification* translates into different delays at different frequencies, dispersing the original signal. In some cases this effect is quite audible. If one uses a *digital approach*, one can either mimic the analog behavior, or use a *linear phase*, a.k.a. *constant delay* filter. This filter will *equally precede and follow the signal*; part of the filter may create a pre-echo effect, modifying the leading edge of transients and signal changes. A high Q linear phase filter can introduce audible pre-echo in the short millisecond range; it's exactly like a floor bounce but without the comb filtering. Any time that a high Q filter is used, careful listening with both types of equalization may be necessary to decide which choice is best.[10]



Phase graph of a 100 Hz parametric boost. In green, linear phase, in red, minimum phase.

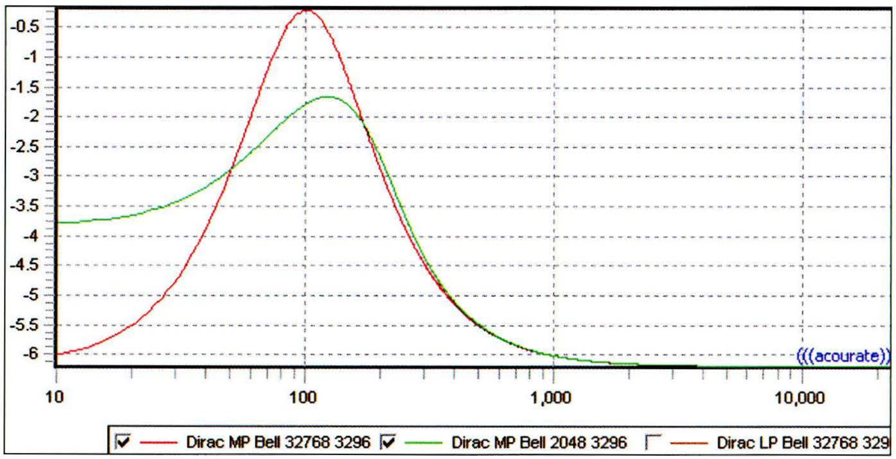


Impulse response of 100 Hz bell curve. In green, minimum phase, in red, linear phase.

This graph (above) plots the impulse responses of these two styles of EQ, made with the help of the excellent analysis program *Acourate*. The horizontal scale is time and the vertical scale is amplitude with each impulse normalized to an amplitude of 1. The pulses are intentionally offset from one another for clarity. A single-sample, positive-going impulse with no EQ would look like a vertical line perpendicular to the 0 axis. But we can see that both equalizers spread the sound. The minimum phase (green) spreads the sound after the impulse. The linear phase (red) spreads the sound equally on both sides of the impulse. Some say that echoing before the impulse should sound unnatural, but please consider the amplitudes. This LP EQ is very well-behaved: its echoes have half the amplitude of the MP and shorter duration. The MP has twice the undershoot (6 dB more), plus an overshoot, and it oscillates twice as long as the LP after the impulse. Some of the LP's pre-echo will be masked by **temporal masking**, as will the post-echo of both impulses. Who is to say which one of these EQs sounds better, based on the appearance of the impulse? I know for sure that they sound very different.

As computer power has improved, it is now possible to implement high quality **FIR** (finite impulse response) filters with a very clean impulse, which can be made either as LP or MP. The DMG Equilibrium is the first plug-in equalizer to give users the choice of phase response, FIR impulse-response length as well as the window parameters. One band of this EQ can be LP, another can be MP: this proves very useful in mastering, as we shall see. If the windowing is wrong, an equalizer can produce ripples in the frequency response, or distortion due to wraparound. The latter is revealed in a clever graph in the DMG setup window. If the impulse response is too short, the frequency response becomes very ragged. So, the longer the impulse length, the more accurately the filter reflects its design goals. The penalty of using long impulse length is latency. This is not a problem in mastering if the DAW has latency compensation, which synchronizes the waveform with the sound being monitored. Latency is of course a problem in recording and mixing, so long impulse response equalizers are not very common. But they do sound better! At lower left is the shape of our 100 Hz bell curve comparing a 2048-sample impulse response

In red, amplitude response of a 100 Hz bell curve made with a 32768-sample IR. In green, the same EQ with only 2048-sample IR.



(green) with a 32768 sample response (red). As you can see, the shorter IR makes a very poor bell curve: it's off frequency, its peak is almost 3 dB low, and it hardly dips at all below 100 Hz. Higher sample rates produce worst discrepancies unless the impulse length is made high enough. High pass filters are very ineffective when implemented with short IR lengths. For mastering and other critical applications I recommend 32768 – or else stick with good ol' fashioned **IIR** (infinite impulse response) filters that do not need an impulse response specification, but could oscillate or cause aliasing at high frequencies.

Linear phase EQ: The Sound

Many of the qualities (including the "bite") we've grown accustomed to in minimum phase equalizers are due to their phase shift. In fact, John Watkinson believes that much of the audible difference between EQs comes down to their different phase response.[11] LP equalizers have come a long way since we published the second edition of this book.

The LP can boost or cut frequencies without altering distance, which in itself is a powerful characteristic if you don't want to alter the depth. During mixing, equalization is usually done on an individual instrument, so even if the instrument moves closer in the soundstage when you boost its frequencies, the mix engineer can control depth with level, reverb, or other tools. But during mastering, it's not good to smear the depth, although we have been doing it for years with our analog equalizers. Altering the level of an LP frequency does not move instruments forward or backward in the soundstage. It's possible to raise or lower the high frequencies without moving the cymbals forward or backward, thus preserving the accurate position of the drumset in the ensemble while avoiding shrinking the distance between the front and back row of the group. So LP is far more useful for mastering whole groups than mixing individual elements. I performed a shootout of the DMG's LP versus MP using a recording of an acoustic ensemble which needed a high frequency shelving boost. The LP was the clear winner, with great depth and a distinct soundstage. The MP appeared to shrink both the width and the depth. To my ears, the LP was as transparent as the MP when the impulse response was set to 32768. There was no loss of clarity, and the LP EQ did not cause any softening of the sound, contrary to my previous experience with LP equalizers. On the contrary, it was the MP equalizer that distorted and warped the sound. And this was a multimiked acoustic recording, not some special "audiophile" production. For me it was no contest. I've already started using LP high frequency shelves in my mastering work.

I especially love the ability of LP to keep high hats from moving forward when boosting in the cymbal range. Alan Silverman (in correspondence) says,

> EQ'ing frequency ranges [in LP] is more like raising or lowering a fader in a mix than EQ'ing.

If you're looking for an aggressive sound, perhaps the MP is your best choice. But if you are looking for a natural sound, particularly at high frequencies, LP is your choice. Narrow-band peaks and dips can be accomplished in linear phase, avoiding the smeary quality that occurs in minimum phase with sharp bands. The jury is still out on which type of equalizer is best for low frequencies. Preliminary comparisons seem to show that LP may sound worse than

"Remember the yin and the yang: Contrasting ranges have an interactive effect."

MP at low frequencies. There is much to learn, and I still have to become familiar with all the advantages and disadvantages of the two types. But it's a brave new world: equalizers have evolved, and it is time for mastering engineers to consider the use of LP equalizers.

Dynamic Equalization

Dynamic equalizers (like the Weiss EQ1-Dyn) emphasize or cut frequency ranges *dynamically*, as opposed to *static*, or *fixed* EQ. Thresholds set a level above or below which a band is dynamically boosted or cut. This extra amount is added to a *static* setting. For example, above the threshold we can lower the high frequency response; we could start a static band at, say +1 dB, but to prevent harshness at high levels, slowly cut the band's level when the signal exceeds the threshold. Dynamic EQs can be used as noise or hiss gates, rumble filters that function at low levels (especially useful for traffic control in a location classical recording), sibilance controllers, presence enhancers or ambience enhancers. They can enhance inner details or clarity of high frequencies at low levels, where details are often masked by noise. Or they can enhance warmth by raising a lower midrange band at low levels, but prevent the sound from getting muddy at high levels. Multiband dynamics processors (see Chapters 6 and 7) can also perform dynamic equalization.

That brings us to the end of our discussion of equalization techniques. The many exciting new equalization tools and techniques we've described here offer lots of new options for your mastering workflow. But keep in mind what we said at the beginning of the chapter: changing anything affects everything. It's just part of the "Law of Unintended Consequences!"

1
Chapter 9 will discuss ways to overcome the first principle, and reduce the compromise.

2
We may believe we have "the absolute sound" in our heads, but are surprised to learn how much the ear/brain accommodates. If we play a bright album immediately after a dark one, at first there is ear shock, but we quickly adapt, though the new sound continues to affect subliminally. The same thing happens in photography and motion pictures, after an initial shock, the change to "Kodachrome" becomes subliminal.

3
Overly bright records become dull on FM radio due to high frequency FM broadcast limiters. Thus FM radio processing makes creating bright recordings self-defeating.

4
In 1967, George Massenburg began the search for a circuit he could use to independently adjust an equalizer's gain, bandwidth and frequency. The key word is *independent*, for most analog circuits fail in this regard and the frequency, Q, and gain controls interact with each other. His circuit, which he calls a **parametric equalizer**, remains proprietary today.

5
Some equalizers define bandwidth in octaves instead of Q. An online converter between Q and bandwidth can be found in the links.

6
With apologies to Captain Kirk.

7
This is dictated by the psychoacoustic *Equal Loudness Curves*, first researched by Fletcher, Harvey and Munson in the 1930's. And revised in ISO 226:2003.

8
Historically, the high pass filter was crucial when making LPs, to prevent excess groove excursion and obtain more time per LP side, but digital media do not have this physical problem.

9
No, Virginia, there is no such thing as an *absence filter*.

10
Jim Johnston, in correspondence.

11
(9/1997) *Studio Sound Magazine*.

"The perfect mix may need no mastering processing at all!"


Allianz

CHAPTER 5

How To Manipulate Dynamic Range for Fun and Profit: Macrodynamics, Loudness Range

I. The Art of Dynamic Range

EBU R-128 Loudness Range

This is the first Chapter in a trilogy about *dynamic range*. The term **dynamic range** refers to the *difference* between the loudest and softest passages of a recording; it should not be confused with **loudness** or a program's average level. Ironically, it wasn't until the year 2012 that we officially agreed on how to measure dynamic range. Before that, we had no measurement method to deal with the following issues:

- If a song fades out to silence, can we claim it has 90 dB of dynamic range?
- Does adding a spoken-word introduction to a hard rock song give it 40 dB of dynamic range instead of only 5?
- Does a 10-second soft passage in a pop song negate the effects of its otherwise constant loudness?

The answer to all three questions is, "of course not." So how can we judge the effectiveness of a brief soft passage in the middle of a highly compressed song? The answers lie in two groundbreaking developments. The first is the international audio standard for measuring loudness, **ITU BS.1770-3** (third version), which tells us how to measure both a program's **Integrated Loudness** (also known as **Program Loudness**) and its **True Peak Level**. It's set by the International Telecommunication Union (ITU). The second breakthrough is the European Broadcasting Union's recommendation **R-128**, which defines a program's **Loudness Range** (abbreviated **LRA**). This is the first formal definition for how to measure dynamic range. In this book, whenever we say "dynamic range," we always mean the EBU-R128 definition *LRA*. In Chapter 16 we will describe in detail how LRA is measured, but for this chapter we'll concentrate on the subjective art of manipulating dynamic range.

A musical work that includes soft and loud passages usually sounds more natural and can sound more exciting than one that does not. The typical measured LRA of a popular music recording is only 6 to 8 dB, but when the music is suitable and I'm given the opportunity, I enjoy working with material that has far greater LRA. In 2012, Rudolph Ortner[1], who was a masters student in audio at the time, produced a comprehensive thesis that measured and statistically analyzed over 10,000 charting popular music recordings made between 1951 and 2011. Ortner determined that the median LRA of popular music has consistently been 6 dB (plus 1, minus 2 dB) for each of the past 60 years! This means that popular music producers have been very consistent in the dynamic range that they judge to be suitable. But it does not mean that there was no loudness race; on the contrary, other statistics uncovered by Ortner reveal the egregious extent of the race. Furthermore, Ortner determined that producers have held soft passages of popular music about 5-7 dB below the program level with some variances (See Chapter 17). Classical, jazz, many acoustic and some electric styles exhibit larger dynamic range. A symphonic recording may have an LRA of 10 to 15 dB, or in rare cases, more. Pictured (below left) is the waveform of a classical recording I mastered with extreme dynamic range, with a very rare LRA of 19.3 dB (pretty much the difference between a whisper and a shout). Reflecting the high standards of its producers, this recording's soft passages strain the ability of even a quiet room — its loud passages entertain the soul and move the gut.

My master of Grand Fantasie from Die Walkure, *performed by the U.S. Marine Corps band, has an amazing extreme LRA of 19.3 dB and its program loudness is -20.5 LUFS.*

Macrodynamics and Microdynamics

Dynamics can be divided further into two categories: **Macrodynamics** — loudness differences between sections of a song or song-cycle, measured by LRA; and **Microdynamics** — the music's rhythmic expression, transient quality, integrity or bounce, which involves the music's short-term peaks. Dynamics processors (such as compressors, expanders) can affect the music's *microdynamics* as well as its *macrodynamics*. Manual gain riding can only affect the music's *macrodynamics*, since we can't move a fader fast enough to affect the short term peaks. But we can affect *microdynamics* by editing very short moments. The micro- and macro- manipulations work hand in hand, and many good compositions incorporate both microdynamic changes (e.g., percussive hits or instantaneous changes) as well as macrodynamic (e.g., crescendos and decrescendos). Think of a music album as a four-course meal: The progression from soup to appetizer to main course and dessert is the macrodynamics. The spicy impact of each morsel is the microdynamics. In this Chapter, we'll concentrate on macrodynamics.

The Art of Reducing (compressing) Dynamic Range

The dynamics of a song or song cycle are critical to creative musicians and composers. As engineers, our quality reference should be the sound of a live performance: we should be able to tell by listening if a recording will be helped or hurt by modifying its dynamics. Even when mastering largely-constructed genres like hip hop, I use the dynamics of live performance as my standard. In a natural performance, the choruses should sound louder than verses, ensembles louder than soloists, and the climax meaningfully louder than the rest. Many recordings have already gone through several stages of compression of dynamic range, and indiscriminate or further dynamic reduction can easily push the clarity and the impact downhill. However, usually the recording medium and intended listening environment simply cannot keep up with the full dynamic range of real life, so the mastering engineer is often called upon to raise the level of soft passages, and/or reduce loud passages, which can be done by **manual compression,**[2] moving the fader up or down, or manipulating gain in a workstation. We may reduce dynamic range (compress) when the original range is too large for the typical home environment, or to help make the mix sound more exciting, fatter, more coherent, to bring out inner details, or to even out dynamic changes within a song if they sound excessive, which is definitely a subjective judgment.

Experience tells us when a passage is too soft. As we mentioned in Chapter 3, the ear's sensitivity changes with the context, so a soft introduction located immediately after a loud song may have to be raised — but a similar soft passage in the middle of a song may be just fine. Meter readings are fairly useless in this regard.

How soft is too soft? As a case in point, the engineers at Lucasfilm discovered that having a calibrated monitor gain and a dubbing stage with a very low noise floor do not guarantee that a film mix will translate to the theatre. During theatre test screenings, some very delicate dialogue scenes were being "eaten up" by the air conditioning rumble and audience noise in a real theater. So they created a calibrated noise generator, labeled "popcorn noise," which could be turned on and added to the monitor mix whenever they wanted to check a particularly soft passage. For similar purposes, our alternate listening room at Digital Domain has a ceiling fan and other noisemakers. Whenever I have a concern, I start the DAW playing a loud passage just before the soft one, and take a walk to the noisy listening room.

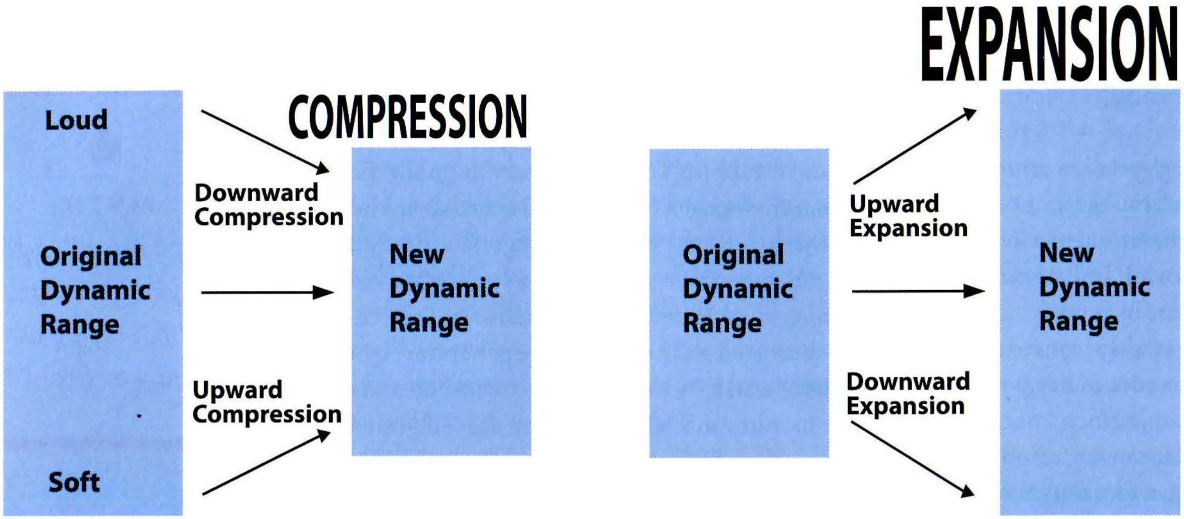
The Art of Increasing Dynamic Range

Increasing dynamic range can also make a song sound more exciting, by using contrast or by increasing the intensity of a peak. The key to success here is to recognize when an enhancement has become a defect — musical interest can be enhanced by variety, but too much variety is just as bad as too much similarity. Passages that are too loud compared to the average can disturb listeners, especially those playing music quietly as background (which, sadly, seems to be becoming the norm). Another reason to increase dynamic range is to restore, or attempt to restore, the excitement of dynamics that were lost due to multiple generations of compression or tape saturation.

The Four Varieties of Dynamic Range Modification

We always use the term **Compression** for the reduction of dynamic range and **Expansion** for its increase. There are two varieties of each: **downward compres-**

MYTH:
"Of course I've got dynamic range. I'm playing as loudly as I can!"[5]



Any combination of these four processes may be employed in a mastering session.

sion, upward compression, downward expansion, and **upward expansion,** as illustrated above.

Downward compression is the most popular form of dynamic modification. It brings high-level passages down. Limiting is a special case: it is downward compression with a very high ratio (*ratio* and other dynamics terms are explained in the next chapter). Examples include just about every compressor or limiter you have ever used. Downward compression can easily be done manually, by simply lowering the level of loud passages, without introducing the artifacts of compression processors. Note that compression processors can decrease microdynamics, while manual compression usually does not.

Upward compression raises the level of low passages. This too can be done manually or through a processor — for example, the **AGC**, which some broadcasters use to make soft things louder. It's the type of compressor frequently used in consumer camcorders, whose pumping and breathing artifacts have given AGC a bad name. In Chapter 7 we will introduce a more effective upward compression processor that is extremely transparent to the ear. For clarity, we will always use the short-term *compressor* to mean *downward compressor* unless we need to distinguish it from *upward compressor*.

Downward expansion is the most commonly used type of expansion: it brings low-level passages down further. Most downward expanders are processors employed to reduce noise, hiss, or leakage. A dedicated noise gate is a special case – that is, downward expansion with a very high ratio. Examples of downward expanders include the classic Kepex and Drawmer gates, Dolby and similar noise reduction systems in playback mode, expander functions in multi-function boxes (e.g. TC Finalizer), and the gates on recording consoles. For clarity, we will use the simple term *expander* to mean the downward type, unless we need to distinguish it from the upward type.

Upward expansion takes high-level passages and brings them up even further. This can be done manually, thereby increasing macrodynamics, or with a processor called an *upward expander*, which can increase both macro- and micro-dynamics. Upward expanders are relatively rare; in skilled hands they can be used to enhance dynamics, increase musical excitement, or restore lost dynamics. Examples include processes available in DMG, Flux, Maselec, UAD, Waves and Weiss dynamics processors. I've spent a lot of time and effort over the past decade working with manufacturers to encourage the development of upward expanders. I'm pleased to see they have begun to proliferate.

II. The Art of Manual Gain-Riding: Macrodynamic Manipulation

In General

During mixing it is difficult to simultaneously pay attention to the internal balances *and* the dynamic movement of the music from section to section — for example, verse and chorus. Sometimes engineers inadvertently lower the master fader during the mix to keep it from overloading. If performed during a build, this will strip the climax of its impact. In mastering we can enhance a well-balanced rock or pop mix by taking the dynamic movement of the music where it would like to go. Delicate level changes can make a big difference: it's amazing what a single decibel can accomplish. It's also important to make sure the client's own level change was not intentional before attempting a correction.

The Art of Changing Internal Levels of a Song: How and When to Move the Fader

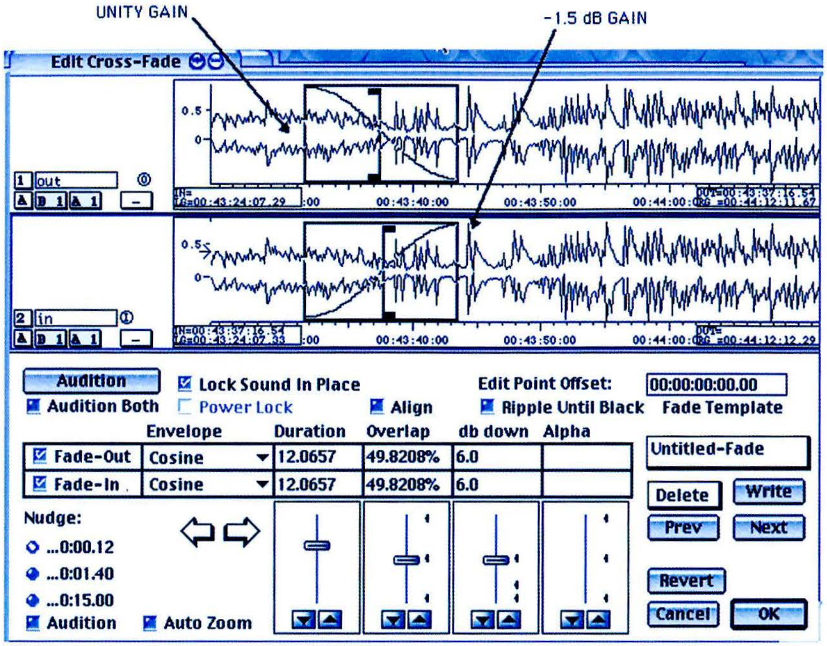
Artistic level changes can really improve a production, but they need to be made in the most musical way. To this end, internal level changes are least intrusive when performed manually (by raising or lowering the fader), as little as a 1/4 dB at a time, as opposed to using processors such as compressors or expanders, which tend to expose their action.

When riding the gain, aim just to augment the natural dynamic flow: if the musicians are trying for upward impact, pulling the fader back during a crescendo can be detrimental, since it will diminish the intended impact. Extra-soft passages require special attention. If the highest point in the song sounds "just right" after processing, but the intro sounds too soft, it's best to simply raise the intro, finding just the right way to restore the gain using one or more of these approaches:

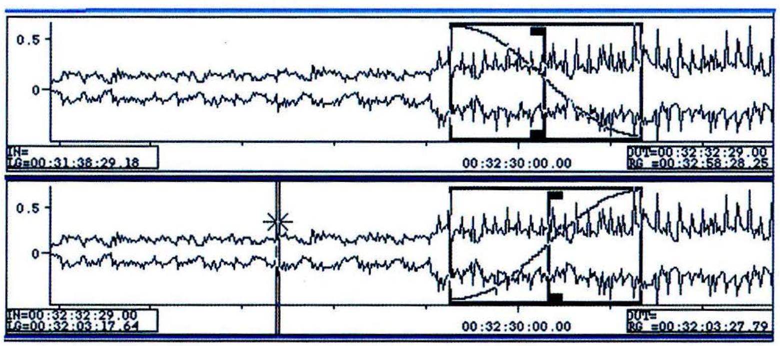
- a quick edit and level change at the transition between the raised-level intro and the normal-level body. This can have a nice effect and be the least intrusive.
- If that doesn't sound good, try a long, gradual lowering of the gain, which might occur at the end of the intro, or slowly during the first verse of the body.
- If that doesn't work, then after the raised intro, try a series of 1/4 or 1/2 dB downward edits, taking the sound down step-by-step at critical moments. This is useful when we don't want the listener to notice that we're cheating the gain back down, and we may be forced to work against the natural dynamics.

Retaining Dynamic Impact while Reducing Dynamic Range

Some soft passages must be raised, but if the musicians are trying to play something delicately, pushing the fader too far can ruin the effect. The art is to know how far to raise it without losing the feeling of being soft, and to find the ideal speed to move the fader without being noticed. In a DAW, physical fader moves are replaced by crossfades, or by drawing gain changes on an automation curve. The mastering engineer's aim is to be invisible; if the sound is being audibly manipulated, the job has not been done properly.[3] Many years ago, I learned a technique for decreasing dynamic range in the least damaging way from Alec Nesbitt's book *The Technique of the Sound Studio*. If we have to take a loud passage down, the best place to lower the gain is at the *end of the preceding soft passage before the loud part begins*.



The modern version of fader-riding. In Sonic Solutions' "classic" edit window the outgoing edit is on top, the incoming on the bottom. Note that the gain drop is performed in the soft passage preceding the loud downbeat, thus preserving the apparent impact of the downbeat.



A soft introduction has been reduced even further, and the impact of the body of the song is enhanced by gradually increasing the gain during the beginning of the main part of the song.

Look for a natural dip or decrease in energy, and apply the gain drop during the end of the soft passage before the crescendo into the loud part begins, or in the gap just before the loud part. In other words, take down the level during a decrescendo, not during the loud passage that follows. That way, the loud passage will not lose its comparative impact, for the ear judges loud passages in the context of the soft ones.

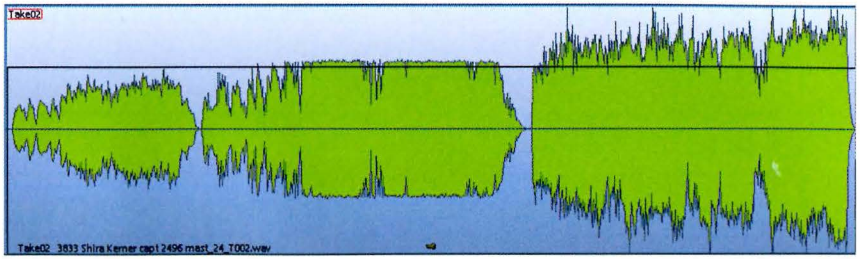
The figure (top left), from a Sonic Solutions workstation, illustrates the technique. The gain change is accomplished through a crossfade from one gain to another. The producer and I decided that the *shout chorus* of this jazz piece was a bit overplayed and had to be brought down from triple to double forte (which amounted to one dB or so).[4] To retain the impact of the chorus, we slowly dropped the level during the soft passage just before the drum hit announcing the chorus. In the crossfade window, we constructed a 12 second crossfade from unity gain (top panel) to -1.5 dB gain (bottom panel); the drum hit is just to the right of the crossfade box. This drum hit retains its impact by contrast, because the musicians' prior delicate decrescendo has been enhanced during mastering.

Enhancing Dynamic Impact in a Natural Way

In this next figure (bottom left) we have purposely lowered the level of an introduction in order to create an impacting crescendo in the body of the song. I reduced the intro and slowly introduced a crescendo (20 seconds long) that enhances the natural build as it goes into the first chorus. The top panel is at -1 dB gain, and the bottom panel is at unity (0 dB) gain, achieved at the end of the crossfade. Another way to increase the dynamic impact is to increase the space between two songs. This extends the tension caused by silence and acclimates the ear to the silence, so the next song's loudness becomes a surprise.

Leveling The Album to Achieve Good Dynamic Range

Here's an example of how I may level the songs in a rock album to achieve good dynamic range. Take a look at the image (top right). The client wants this acoustic rock album to be mastered "as open and natural and dynamic as possible." The second track was mixed by a different mix engineer, who peak-limited and squashed his mix. I improved its sound, helping it to sound more open, through micro- and macro- dynamic techniques, but you can still see that the peaks of this song are totally gone. The important thing to learn is that *waveform peaks have nothing to do with a program's loudness*. For example, the second song looks like a flat square wave. The first and third songs, however, have considerable short-term peaks, which we need to preserve if we want the album to sound open. It may surprise you to learn that the second and third track are equally loud! The first track pictured is a ballad and so it has to be made soft. The third track sounds much more open, spacious and dynamic, and as you can see, uses up all of the available headroom with its superior peak-to-loudness ratio, which we'll discuss in the next chapter. To master this album, I would start by mastering the third track, taking it to its sonic potential, and adjusting the loudness of all the other tracks relative to it. If I had started mastering with the second track, and had made it louder, it would have forced me to squash the peaks of the third track, taking away its "open" quality. So the best thing to do is to *leave headroom*, and not fill up all the peak space with the limited tune.



An album that begins with a soft ballad, followed by a compressed rocker, followed by a more open-sounding rocker. The second and third tunes are equally loud!

In Conclusion

Macrodynamic manipulation is a sometimes overlooked but powerful tool in the mastering engineer's arsenal. In the next installment of our Dynamics Trilogy we move on to the use of compressors, expanders and limiters to manipulate both macro- and micro- dynamics.

1
Ortner, Matthias Rudolf (2012) Je lauter desto bumm! (The Evolution of Loud). Master's thesis, Danube University Krems

2
Please do not confuse the term *dynamic range reduction* (compression) with *data rate reduction*. Digital Coding systems employ data rate reduction, so that the bitrate (measured in kilobits per second) is less. Examples include mp3, AAC, Dolby Digital (AC-3). Since it's not good to refer to two different concepts with the same word, we should encourage people to use the term *Data Reduction System* or *Coding System* when referring to data. Use *compression* only when referring to the reduction of dynamic range.

3
This is true for most of the "natural" music genres, with some exceptions being hip-hop, psychedelic rock, performance art, etc. where the artists invite the engineer to contribute surprising or rococo dynamic effects.

4
Producers don't always use classical Italian dynamic terms to describe their needs. The mastering engineer should choose the bonding language which is best for the client – "Make it louder, man!"

5
A common misconception. Thanks to Gordon Reid of Cedar for contributing this audio myth.

"Waveform peaks have nothing to do with a program's loudness."

CHAPTER 6

How to Manipulate Dynamic Range for Fun and Profit: Downward Processors

I. Introduction

Chapters 6 and 7 discuss those ubiquitous devices we call *dynamics processors*. We must study their objective characteristics to learn how to use them effectively. In this Chapter we begin with traditional compressors and limiters, the vast majority of which are "downward processors," to see how they affect the *macrodynamics* of sound and musical movement — in other words, the *variations in the loudness of the music*. As we move along in the Chapter we'll also see that downward processors can affect music's *microdynamics* — that is, the momentary (short) transients, that affect the perceived quality and clarity of the sound.

If I raise the gain of a recording by turning up a fader, it will obviously sound louder. And its loudness goes up linearly with the movement of the fader: for every dB that the fader is raised, the loudness goes up one dB, or more formally, one **LU** (loudness unit). But if that recording's peak level is already at full scale, seen as 0 **dBFS** on a peak meter, raising the fader will *clip*, or *overload* the sound. It may still sound somewhat louder, but that additional loudness is now accompanied by the distortion of the digital medium. From this point on, as I raise the fader, the perceived loudness increase is no longer proportional to the amount that I raised the gain, because the digital medium itself pulls back the peaks of the sound while it adds distortion. By clipping, the medium itself is *peak-limiting* the signal, in the most primitive and distortion-ridden way. The first compressors and

"Downward compression makes the loud passages softer."

limiters were invented in the analog era to prevent overload without causing clipping distortion — specifically, to reduce the saturation of analog tape or prevent the overmodulation of a radio transmission. Afterwards, audio engineers discovered their use for creative purposes: to modify sound, to help create **punch**, to aid in mixing because they allowed softer instruments to compete effectively in a mix with louder ones, and many other creative uses.

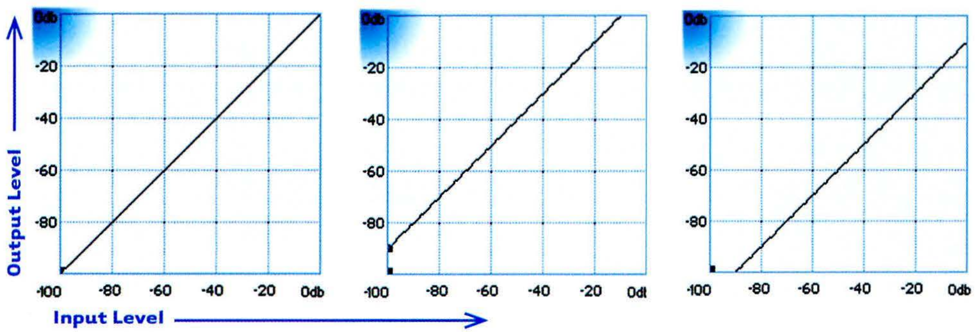
II. Compressors and Limiters: Objective Characteristics

Transfer Curves (Compressors and Limiters)

Linear means a straight line. A transfer curve (or plot) displays the input-to-output gain characteristic of an amplifier or processor. Input level is plotted on the X axis, and output on the Y. A *linear* amplifier shows a straight line (not a curve), hence the name. The figures (below) show a family of linear plots at 3 different gains. **Unity gain** means the ratio of output to input level is 1, or 0 dB, so a unity-gain amplifier shows a straight diagonal line across the middle at 45°, called the *unity gain line*. From left to right: unity gain; 10 dB gain; 10 dB attenuation. Notice that the middle plot would clip the medium, and would yield distortion for any input signals above -10 dBFS.

The **threshold** of a compressor is the level at which gain reduction begins. **Compression ratio** describes the relationship between input and output above the threshold. Figure A (next page) is a simple compressor with a fairly gentle 2.5:1 compression ratio, and a threshold at around -40 dBFS (which is quite low and would yield strong compression for loud signals). 2.5:1 means that an increase in the input of 2.5 dB will yield an increase in the output of only 1 dB, or for an input rise of 5 dB, the output will rise by only 2 dB, or as can be seen in the plot, an input change of 20 dB yields an output change of slightly less than 10 dB (once the curve has reached its maximum slope). Notice that *downward compression always makes the loud passages softer*, because above the threshold the output is less than the input.

In Figure B (next page), we add gain after compression so that a full level (0 dBFS) signal input will yield a full level signal output. This control on compressors is often called **gain makeup** (a simple gain amplifier after the compression section), allowing the average level of the material to be raised, while the loudest passages are still brought down. This is the essential key to the action of downward compressors: while they bring down the loud passages, the middle passages can be raised, making the average level louder (because the music spends most

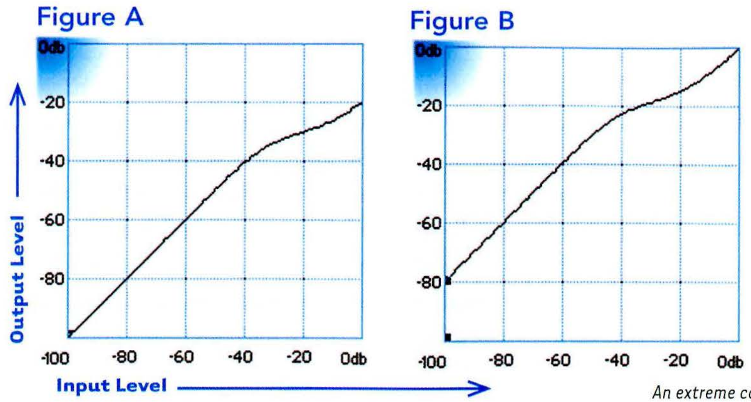


Three transfer curves. *(From left to right). Amplifiers with Unity-Gain, 10 dB gain, and 10 dB loss (attenuation).*

of its time at the middle levels). It is the *gain makeup* that makes the output of a compressor appear louder, though there are other factors that come into play, such as release time, to be explained shortly.

For illustration, we show an amplifier with an extreme amount of gain, 20 dB, which would considerably amplify soft passages (below the threshold). In typical use during mastering, however, makeup gains are rarely more than 1 to 3 dB. If there is no gain makeup, the output of a compressor will sound softer than its input. We can take advantage of that during mixing: the compressor softens the loud moments of an instrument. Then we can bring its level down but not lose its soft passages within the mix, since it now has reduced dynamic range. The mix fader for each instrument during mixing effectively acts as another post-compressor gain control. Or if you have placed a compressor on each track, you could set all the faders in a line at 0 dB and control the mix levels with each compressor's gain makeup knob. I'm not advising putting compressors on every track (in fact, I advise against it), I'm just demonstrating that knowing what's under the hood is as important as what's visible on top.

In this extreme example loud input passages from about -40 to about -15 are still amplified. Above about -15 dBFS, the curve slopes back to unity gain, reduces the total gain, and resembles that of a linear amplifier. Far below the threshold, it's a fairly linear 20 dB amplifier and can have fairly low distortion, because there is no gain reduction action. The action of the compressor that most affects the sound quality occurs in the area where the line is curved. When the engineer wants to have the most sonic effect, he seeks to place the music levels in the curved area. In the illustration, that would



An extreme compressor for illustration: Figure A, 2.5:1 ratio, -40 dBFS threshold and no gain makeup. Figure B, the same compressor with 20 dB gain makeup.

correspond to input levels between about -40 and -15 dBFS. This can be done by adjusting the threshold until some gain reduction action is seen during the major passages of the music. To get the greatest esthetic effect from any compressor, most of the music action must occur where the curve's shape is changing. Thus, a real-world compressor's threshold would likely be -20 to -10 dBFS or higher. At full scale in Figure B, 20 dB of gain makeup is summed with 20 dB of gain reduction, yielding 0 dB total gain. This compressor model's curve levels off towards a straight line above a certain amount of compression, so the ratio holds true only for the first 15-20 dB above the threshold. Other compressor models continue their steep slope, thus maintaining their ratio far above the threshold. There are as many varieties of compression shapes as there are brands of compressors, and they all produce different sounds.

Never Trust A Gain Reduction Meter!

The **gain reduction** (GR) meter in a compressor tells us when the signal has exceeded threshold and how much the compressor is reducing the gain of the signal. These meters vary in accuracy, both in digital and analog units. In many cases, the compressor is found to be doing something even when the meter is not moving.

In some analog units GR meters are much slower acting than the actual gain reduction, so it is possible to have 3 to 6 dB of real gain reduction while the meter shows only 1 or 2. On the other hand, some analog units use a fast, peak-sensitive GR meter, which does show us the action. But peak-sensitive GR meters respond faster than the ear can detect, and so give us a falsely high indication. In the case of digital units, the programmer rightly spends more of his time making the compressor work than updating the GR meter, so the meter may appear sluggish — especially if there are many instances in use in a single DAW. I advise caution! Learn how the GR works in each model compressor you use. But even then, mostly ignore your eyes: use your ears and never forget the facts of this paragraph!

Figure A

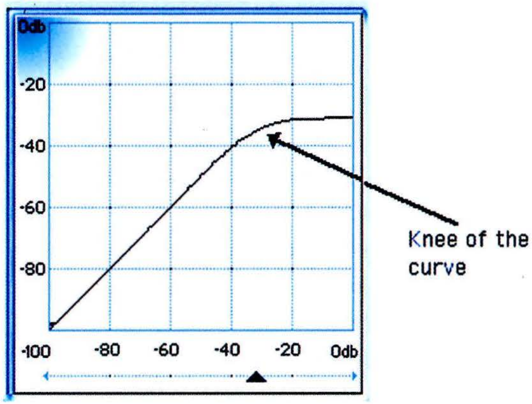
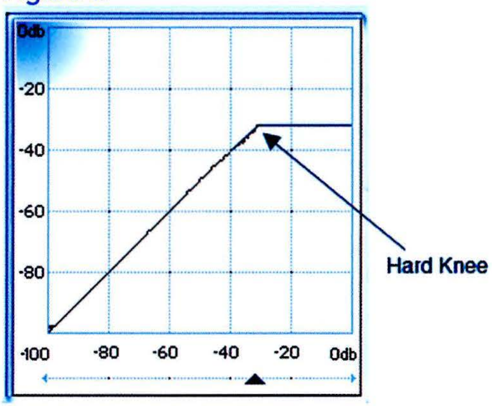


Figure B



Knee shapes: Figure A, compressor with soft knee. Figure B, hard knee.

Knee Shape, Compressors vs. Limiters

Figure B (this page) shows a very high ratio of 10:1. Above the threshold, the output is a horizontal line, which is very severe compression, commonly called **limiting**. The definition of limiting is really a matter of degree, but most authorities call a compressor with a ratio of 10:1 or greater a **limiter**. In other words, a low ratio device is usually called a compressor, and a high ratio device is called a limiter. Very few analog limiters have ratios higher than 10:1. However, some digital limiters have ratios of 1000:1 to prevent the slightest overload. The portion of the curve near the threshold is called the **knee**, which marks the transition between unity gain and compressed output. In limiters, the knee should be very sharp (also referred to as "hard"). Compressors can have hard or soft knees. **Soft knee** refers to a rounded knee shape, or gentle transition (Figure A); **hard knee** refers to a sharper shape (Figure B), where the compression reaches full ratio immediately above threshold. Soft knee can sweeten the sound of a compressor near the threshold. For those models of compressors that have only hard knees, some of the effect of a soft knee can be simulated by reducing the ratio or raising the threshold, which will result in less action.

Attack and Release Times

Attack time is the time it takes for a compressor to implement full gain reduction after the signal has crossed the threshold. Because digital compressors can react with essentially infinite speed, a digital compressor set to 100 ms may sound similar to an analog compressor set to, say, 40 ms. The method the designer uses to define attack time is not standardized, so it's not possible to compare specified attack times between brands. It's better to remove all the labels (except *slow* and *fast*) and just listen! With digital compressors, typical attack times used in music mastering range from 30 ms to 300 ms (or longer on occasion), with the average time used being probably 100 ms. But go by your ears, not by the numbers. To help you set attack time, listen to the percussive and transient quality of the music: shorter attack times soften transients and produce a more "closed" sound; longer attack times let the music breathe and reveal more of the percussive transients.

Release time, or **recovery time**, is how long it takes for the signal level to return to unity gain after it has dropped below the threshold. Typical release times used in music range from 50 ms to 500 ms, or as much as a second or two, with the average being around 150-250 ms.[1] The terms *short* or *fast* with attack or release time are used interchangeably, as are *slow* and *long* attack and release times. Manufacturers may measure times to 90% of gain reduction, or use another empirical approach to define them. Release time is probably the single most influential setting affecting the "sound" of a compressor. Super-fast release times help make the sound appear unrelentingly loud and aggressive and slow release times are more gentle on the sound. Analog optical compressors have a fast initial release and then a slow final release, which yields a more "gentle" aspect and "bloom" to the sound. VCA compressors produce the reverse effect, which can aid in producing a more aggressive sound quality. Digital compressors attempt to emulate one or the other of these analog characteristics, or be switchable to do either. A good starting point for a digital compressor on mixed music is to set attack time to about 100 ms and release to about 250 ms. Then listen and adjust. If you want a more punchy, aggressive sound, shorten the release slightly and, if useful, shorten the attack from there. Higher ratios, harder knees and greater gain reduction also contribute to a more aggressive sound, but be careful, when a parameter is turned too far, the sound loses its definition and punch — as with any process, often less is more.

With digital limiters, release time is very important: the faster the release time, the more invisible the limiter can be: it jumps in, quickly controls the transient, then gets out of the way. The fastest digital limiters may have a release time of only 1 ms. However, super-fast release times can cause significant distortion. This is why the most successful digital limiters have auto-release control, which slows down the release time if the duration of the limiting is greater than a few milliseconds. The effective release time of an auto-release circuit can be as short as a couple of milliseconds, or as long as 50 to 150 milliseconds. All of this is intended to make the digital limiter as invisible as possible.

"The key to a great master is to start with a great mix."

Release the Automatic Weapons!

Some compressors exhibit automatic, or program-driven release and/or attack times. The designer's choice of automatic settings depends on his idea of what kind of compressor he is creating:

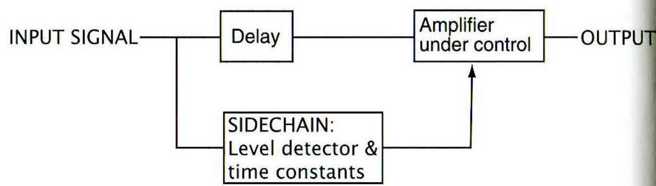
- aggressive
- gentle (invisible)
- distorted (fast release times can cause low frequency distortion, which adds either desirable or undesirable character, depending on the music)
- clean

Some designers try to get the fastest release time possible in automatic mode without causing overt distortion, the goal being to achieve the most aggressive sound. Whether you choose automatic depends on your goals: quick-and-dirty is rarely the goal of the mastering engineer. We prefer to slave over a compressor and optimize the attack and release times until they are perfect for the music in question (although I once discovered a manufacturer's preset which did much more for the music than any attack or release time I could dial in). Manufacturers label their automatic modes in different ways (e.g., **PDR**, which means program-dependent release) or by giving presets. Kudos to Crane Song for implementing a large set of presets that give a variety of character possibilities to their Trakkers and other models. This changes the shape of the attack/release characteristic, but still gives the user the ability to adjust attack and release times, the best of both worlds. *Vari-Mu* analog compressors purposely have gentle attack/release characteristics, which give them their distinctive, sought-after "creamy" sound.

Some compressors have automatic gain makeup, which means that as the threshold is lowered, the gain is automatically raised. Personally, this feature irritates me, but you may like it. I never use it during mastering but I can see its usefulness in mixing. George Massenburg's digital compressor has a feature that psychoacoustically scales the gain with the amount of compression or ratio, so the mix engineer can create the sound character of an instrument he likes while retaining its place in the mix.

Preview, Look-Ahead, Side Chain

The preview, or **look-ahead** function allows very fast, or even instantaneous (zero) attack time, which is especially useful in a peak limiter to prevent overloads. The unit can effectively react to the transient before it has occurred! This requires a delay line, so analog processors do not have look-ahead. When there is look-ahead, the attack time can be as short as we desire, controlling any peak that concerns us. Look-ahead is only relevant when we want short attack times, since if we want a long attack, then we probably also want to let initial transients pass through. So look-ahead is probably unnecessary for attack times longer than about 10 ms. Every compressor has a **sidechain**, which is the control path (as opposed to the audio path), as illustrated in this figure. The compressor places a time delay in the audio signal chain, but not in the sidechain. This gives the sidechain time to "anticipate" the leading edge of a transient and nip it in the bud. The delay



Look-ahead and sidechain in a digital compressor

time only has to be as long as the sum of the shortest transient duration we want to control, plus the reaction time of the sidechain circuit. Analog compressors have certainly gotten along splendidly without preview delay: in fact, much of their sonic virtue comes from their *inability* to stop initial transients. Sharp transients contribute to the life and impact of a recording, so I remove them only when they're audibly objectionable. The exceptions to this might be transients shorter than the ear can hear, which often occur in digital recording, and would prevent the program from having a higher loudness. Removing these was the initial purpose of the brickwall limiter, designed to be short, quick, and invisible.

Figure A

Figure B

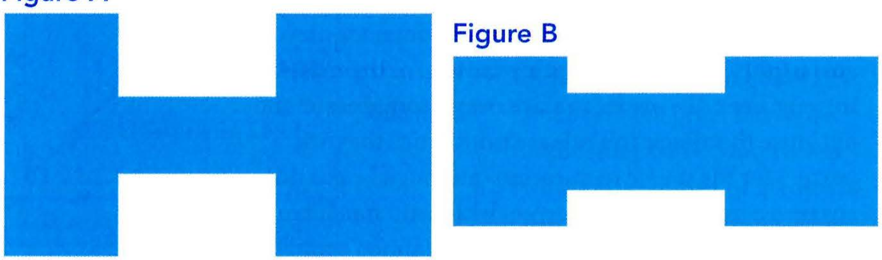


Figure A: a simple tone burst from high to low level and back. Figure B: the same tone burst passed through a compressor with very fast attack, high ratio, and fast release time.

Figure A on page 86 is the envelope shape of a simple tone burst, from a high level to a low one and back again. Figure B is the same tone burst passed through a compressor with a very fast attack, high ratio, and fast release, and whose threshold is midway between the loud and soft signals. Note that the loud passages are instantly brought down, the soft passages are instantly brought up, and there is less total dynamic range, as shown by the relative vertical heights (amplitudes).

Figure C (this page) is the envelope of a compressor with a low ratio, slow attack time, and slow release time. Notice how the slow attack time of the compressor permits some of the original transient energy of the source to remain until the compressor kicks in, at which point the level is brought down. Then, when the signal drops below threshold, it takes a moment (the release time) during which the gain slowly comes back up. A lot of the compression effect (the "sound" of the compressor) occurs during the critical release period, when the gain is coming back up, since except for the attack phase, the compressor has actually reduced the gain of the high level signal.

Compare with Figure D (this page), a compressor with a much higher ratio, faster attack, and very fast release. The higher ratio clamps the high signal down further, and the fast release time aggressively brings the level up as soon as the signal drops below threshold. This type of fast action can make music sound *squashed*, because it quickly brings down loud passages, raises soft ones, and shortens transient attacks.

Figure C

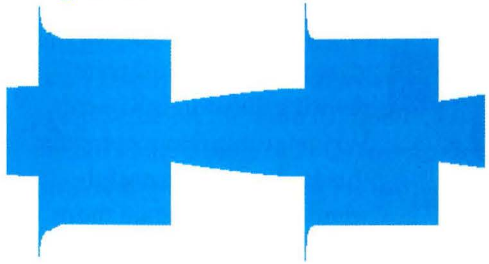


Figure D

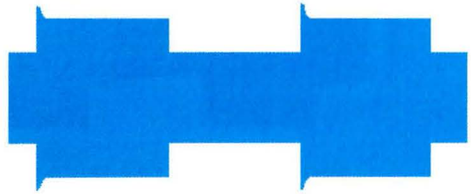
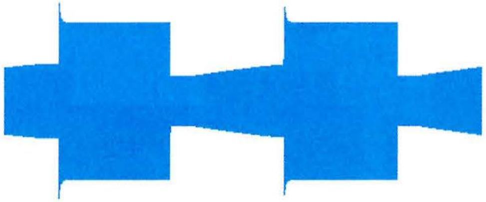


Figure C, a compressor with a low ratio, slow attack time, and slow release time. Figure D, higher ratio, faster attack and very fast release.

The essential fact here is that (downward) compressors take the loudest passages *down*. Gain makeup allows the *average level* to be raised, but the loudest passages end up proportionally lower. A compressor can add or enhance **punch** in mastering, since its essential mechanism is to reduce the partial loudness of peaks and bring up the mid levels; if the compression is overused, or if too many transient attacks are softened, we can lose punch. Remember that *manually* raising the soft passages (avoiding the processor entirely) can leave the loud passages untouched, while giving the production more impact, and preserving the microdynamics that compressors can take away. In the next Chapter, we will investigate how upward compressors can increase soft and mid-level passages with little effect on the important loud ones — a technique that can produce a recording that is dynamic, loud, and has impact at low levels. In Chapter 16, we'll discover a measurement called **PLR**, which can be used as an indication of whether microdynamics are being maintained, or whether they have been lost.



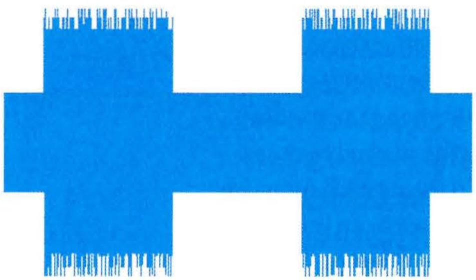
Output of a compressor with a low ratio, slow attack time, slow release time plus release delay.

Release Delay

A release delay control gives us more flexibility in painting the sound character. Very few compressors provide this facility but it is useful when we want to retain more of the natural sound of the instrument(s), and not exaggerate its sustain when the signal instantly goes soft, or reduce "breathing" or hissing effects when the source is noisy (illustrated at left).

Attack/Release Distortion

Part of the sonic aggressiveness of fast release times comes from the distortion at low frequencies that can occur if the release time is too fast. This figure (at left) illustrates what happens when the attack and release times are much too fast. The distortion shown here is caused by the compressor's action being so fast that it follows the shape of the low frequency waveform rather than the overall envelope of the music. This problem can occur with release times shorter than about 50 ms, and correspondingly short attack times.



With attack/release too fast, a compressor can produce severe distortion.

III. Dynamic Manipulation: Adjusting the Impact of Music with a (downward) Compressor

The Engineer as Artist

Compressors, expanders, and limiters form the foundation of modern-day recording, mixing and mastering. With the right device we can make a recording sound more or less percussive, more or less punchy, more or less bouncy — or, put more simply, bad, mediocre, good or excellent.

In skilled hands, compression can help produce a wonderful recording. A skilled engineer may intentionally use creative compression to paint a sound and form new special effects. A lot of contemporary music genres are based on the sound of compression, both in mixing and mastering, from Dance to Rap to Heavy Metal. The key words here though are *intent and skill*. Surprisingly, however, some engineer/artists don't know what uncompressed, natural-sounding audio sounds like. While more and more music is created in the studio control room, it's good to learn how to capture natural sound before moving into the abstract. Picasso was a creative genius, but he approached his art systematically, first mastering the natural plastic arts before moving into his cubist period. Similarly, it's good practice to know the real sound of instruments. Recording a well-balanced group in a good acoustic space with just two mikes is a lot of work, but also a lot of fun! Before multitracking was invented, there was much less need for compression, because close miking exaggerates the natural dynamics of instruments and vocals. At first, compressors were used to control only those instruments whose dynamics were severely altered by close miking, e.g. vocals and acoustic bass. When modern music began to emphasize rhythm, many melody and harmony instruments became masked by the rhythmic energy, inspiring the creative possibilities of compressors and a totally new style of recording and mixing. The advent of the SSL console, with a compressor on every individual channel, changed the sound of recorded music forever.

The Mixing Engineer's Approach to Compression

Let's talk about using compressors in mixing practice. There are at least three styles of mixing:

1) *Move the faders, avoid compressors!* I'm a big fan of retaining transients and impact, so I prefer not to purposely mix into a bus compressor. For the styles of music that I like to mix, I prefer to avoid compressors until one is needed to help bring up the low levels of an instrument that is a bit too dynamic and riding its levels is impractical. Then, if that's not enough, I may put a compressor on a submix bus. It is very easy to become dependent on a bus compressor to "perform the mix" for you, when the traditional method has been to sweat bullets and move faders all day. No one can deny that moving faders is a more pure form of the art, because a compressor not only brings up low instruments, it also changes their sound and their transient characteristics.

2) *An intermediate mixing approach* is to first get the very best mix possible without a bus compressor, then add a bus compressor, and carefully switch it in and out, comparing the sound at equal loudness (don't turn the gain makeup up further than is necessary to match the loudness in bypass). If the bus comp version sounds better than the raw mix, then use it! Sometimes the inner details that the bus comp brings up add interest to the mix, but sometimes the loss of dynamic movement takes away the mix's excitement. This is an esthetic call. I also recommend sending two versions to the mastering engineer, who may have a compressor which is better-tuned to the music and retains impact and transients better than the compressor chosen by the mix engineer.

3) *The more aggressive approach* — relying on a mix bus compressor to perform many of the mix duties — can easily become a crutch that substitutes for good mixing practice. Or, in the hands of a skilled mixing engineer, the bus comp becomes a tool to help deliver the attitude and punch he wants for the particular song.

What about mix program levels when applying a bus comp? In the end it is the sound that counts, not the levels. In other words, you can produce a program whose average level is, for example, -20 dBFS or a program whose average level is -14 dBFS and achieve *the same sound quality* when the two programs are compared at equal loudness. The key is to adjust the compressor's threshold and/or the level of the faders coming into the compressor to achieve the amount of gain reduction you desire. Higher fader level with higher threshold can yield the same sound as lower fader level with lower threshold: just adjust your monitor gain to yield the same loudness and you will see that the results are the same. If the program is to be later mastered, then the mastering engineer will take care of the program loudness. Mix for the sound quality, and don't be fooled if the loudness of your mix is higher or lower than "the competition." This can probably be taken care of later in the mastering. Send a mix to the mastering engineer for evaluation/confirmation.

Compression and Limiting In Mastering

Mastering requires us to develop new skills, since it is concerned with overall mixes rather than individual instruments. Compression is a tool that can change the inner dynamics of music — e.g., by enlivening low- and mid-level passages, enhancing rhythmic

> *"Compression is for kids... it's a crutch."*
> — Bruce Swedien

"*An improperly set compressor can degrade the snap, life, and the punch.*"

movement, or producing a stronger musical message. On the other hand, the intent of a digital limiter is not to change the sound very much, but simply to increase the program level. However, analog limiters fall somewhere in between compressors and digital limiters. They're often used in mixing, but less in mastering, being neither capable of subtleties, nor able to completely fix instantaneous program overloads. As with compressors, it is the gain makeup process that lets us make the output of a limiter louder. In mastering, the most popular compressor is one with a low ratio, as low as 1.5 or 2:1. Regardless of whether we're compressing or limiting, when the peaks have been brought down, there is room to bring the average level up without overloading. For example, snare drum hits that momentarily stick out above the average can be softened by the peak limiter (if this change in sound proves desirable), and the average level can then be raised. That's why digital limiters are used more often in mastering than in mixing. Even the best limiters are not completely inaudible: they can soften the transients and dull the "sharpness" of the sound. As I mentioned in Chapter 2, make sure your ears and monitor system are capable of discerning the sound effect of transient loss, to prevent transient degradation by ignorance (except by intent, since some genres intentionally employ extreme processing that nearly deletes the transients).

BBC research in the 1940s demonstrated that distortion shorter than about 6-10 ms is fairly inaudible, hence the 6 ms integration time of the BBC **PPM** meter (supplanted by modern loudness meters). But this judgment reflected the limitations of 1940s technology. With today's digital recording and solid-state equipment, transient overloads as short as 1 ms will audibly change or distort the sound of the initial transient (particularly noticeable with solo acoustic piano). With good compressors, limiters and mastering technique, program material with a peak-to-average ratio of 18 to 20 dB can often be reduced to about 14 dB with little effect on the clarity of the sound and some subtle reduction in transient impact. That's one of the reasons 30 IPS analog tape is so desirable: it has this limiting function built-in. A rule of thumb is that short duration (a few ms) transients of *unprocessed digital sources* can often be transparently reduced by 2 dB, and in rare cases as much as 6 dB. with little effect on the sound. *However, this cannot be done with analog tape sources,* which have already lost the short-duration transients. Any further transient reduction by compression or limiting, whatever its desirability, will not be transparent. Limiter distortion is especially audible on material which already has little peak information, because a limiter is not designed to work on the RMS portion of the music and it can sound harsh when pushed.

Make it Snappy — or Punchy?

So in general, the less the amount of limiting and the longer the attack time of compressors, the *snappier* the sound will be. **Snap** is the important companion to punch: a good engineer needs to concentrate on both attributes. We could call the sound of the beater of the bass drum its *snap*, and the resonance of its diaphragm and body its *punch*. *Snappy* reflects the presence of transients and microdynamics in the sound – *snap* is very short upward-moving dynamic contrast that is so short-term that it is not perceived directly as a loudness increase, although it contributes to the partial loudness and the liveliness. *Punch* is dynamic contrast that is definitely perceived as a momentary loudness increase, which can be accomplished in mixing—for example, by adding a brief low frequency sound effect underneath a single bass drum beat to increase its duration and power. A popular way to create or enhance punch is to push the level dramatically into a compressor with a fairly-short attack time and fairly-fast release, which increases the sound power and sustain during its release phase. In order to continue to be effective during moments when punch is being applied, the level has to dip just a little each time in order for it to come back up and "hit you": *without a counterswing there can be no swing*. Not everyone is skilled at creating punch: if not well-done, the sound becomes "wimpy loud." I am reminded of theatre trailers for effects movies that are thrown together using a limiter as a hammer, with effects that sound relentlessly loud but have lost their attack, and have absolutely no punch. When I attend the release of that film, I am relieved to discover the film is actually well-mixed and the problem was only in the trailer. In any well-balanced rhythmic recording, it's important to have both short transients as well as power and punch. Many engineers pay little attention to the snap, and concentrate only on the punch. Be sure you are aware of and can identify both.

Keep in mind that *punch* and *impact* have to begin with the arrangement, the performance and the mix. In many cases, punch goes downhill if we master with the goal of making a recording sound louder. We have to work to make sure punch is not lost during mastering. An improperly set compressor can degrade the snap, the life, and the punch. But given a good, clean mix with a lot of life to it, a clever mastering engineer, armed with the right compressor having just the right attack, release, ratio, threshold, and choice of compressor can produce a recording that has more punch than the mix, and retains just the right amount of snap. The converse is also true: it's hard, if not impossible, to make a punchy master from a dirty mix that has been overcompressed or that has extreme distortion. The attempt will fail if the mix itself has no transient impact.

Sometimes producers deliberately make a master with no transients or dynamic movement. Though this is far from natural-sounding, it has become an acceptable sonic genre in alternative rock—to cite just one example. First symptom: the choruses are lower than the verses! It's worth repeating that it is nearly impossible for us to deliver a punchy or snappy master from a squashed mix. I believe this genre is a direct result of a vicious circle, with producers imitating in a mix the sound of squashed masters they have heard. It is coupled with the tempting availability of a compressor on every channel: When you have a hammer, everything looks like a nail. But the producers don't realize that distortion accumulates, and that squashed mixes will sound like mud on the radio. So they have boxed their mastering engineer into a corner. To continue the analogy, they have applied a hammer to pound a giant nail in place when a small tack would have done the job far better. Still, this sonic genre will remain legitimate until dynamics become popular again.

In an ideal mastering session, if a limiter is used at all, it should be acting only on occasional inaudible peaks, or perhaps a bit more if we like the slight softening effect. A manual for a certain digital limiter reads "For best results, start out with a threshold of -6 dBFS." This is like saying "always put a teaspoon of salt and pepper on your food before tasting it." One modern R&B album is so overlim-

"When you have a hammer, everything looks like a nail."

"Many of these sampled kick drums are all boom without the beater."

ited that the bass drum punches a hole in the vocal on every attack; I doubt this is artistically desirable.

It is a common misconception that a limiter is a *peak protection device* for mastering. It may be used as such in radio broadcasting (to protect the transmitter) or sound reinforcement (to protect loudspeakers or handle an unpredictable group), but in mastering (or mixing), the engineer is not doing a live show: we have total control over our levels and we can make the choice of whether to turn them down or raise them and use a peak limiter. We could simply lower the level a dB or more instead of choosing to peak limit. As the loudness wars ease up, there will be less abuse or unnecessary use of digital peak limiters.

The World's Most Transparent Digital Limiter

The most transparent limiter is no limiter at all! If there is a very short peak (transient) overload, for example, a drumbeat within a section which needs to be made louder, a skilled mastering engineer can use the DAW's editor to perform a short-duration gain drop that can be quite inaudible. This **manual limiting** technique lets us raise the program loudness without inducing distortion from a digital limiter, so it is the first process to consider when working with open-sounding music that can be ruined by too much processing. We can often get away with 1 to 3 dB manual limiting typically for a duration of less than 3 ms. But, longer duration manual gain drops will affect the sound as much as or more than a good digital limiter.

Equal-Loudness Comparisons

When it comes to using compressors, loudness has a major effect on judgment, so it is very important to make compression comparisons at equal apparent loudness. (We've seen the same effect in our discussion of equalizers.) If the processed version is played louder than the unprocessed version during an instant A/B comparison, the former may initially seem to sound better, but long-term listeners usually prefer a less fatiguing sound that "breathes." When we compare at matched loudness, we might be surprised to discover that the processing makes the sound worse, and the "improvement" was an illusion. When making an album at "competitive loudness level," it's a relief if the mastering has not degraded the sound of the program, and ecstasy if it has improved it.

The Nitty-Gritty: Compression in Music Mastering

Consider this rhythmic passage, representing a piece of modern pop music, using large and bold text to represent louder sections:

shooby dooby doo **WOP**...
shooby dooby doo **WOP**...
shooby dooby doo **WOP**

The accent point in this rhythm comes on the backbeat (**WOP**), often a snare drum hit. If we strongly compress this music piece, it might change to:

SHOOBY DOOBY DOO WOP...
SHOOBY DOOBY DOO WOP...
SHOOBY DOOBY DOO WOP

This completely removes the accented feel from the music, which is probably counterproductive.

A light amount of compression might accomplish this:

shooby dooby doo **WOP**...
shooby dooby doo **WOP**...
shooby dooby doo **WOP**

This could be just what the doctor ordered, because strengthening the sub-accents may give the music more interest. Unless we're trying for a special effect, or purposely creating an unnatural sound, it's counter-productive to go against the natural dynamics of music (like the TV weatherperson who puts an accent on the wrong syllable because they've been taught to "punch" every sentence: "The weather *for* tomorrow will be cloudy"). Much hip hop music is intentionally unnatural – anything goes, including the eradication of any resemblance to the attacks and decays of real musical instruments. The use of samples that are already compressed is another contribution to the loss of transients in hip hop. Sometimes the increased punch wins, but please note that many of those sampled and pre-compressed kick drums are all boom without the beater!

Typical Ratios and Thresholds

Compression ratios most commonly used in mastering are about 1.5:1 or 2:1, rarely more, even for the most aggressive rock production with a reasonable program loudness. One way to start compressing to help promote *punch* or *attitude* is to first find the threshold, using a very high ratio (say 4:1) and very fast release time (say 100 ms). Then you adjust the threshold until the gain reduction meter bounces as the "syllables" you want to affect pass by, and you hear this bounce. This ensures that the threshold is optimally placed around the musical accents you want to manipulate – the "action point" of the music. Then reduce the ratio to very low (say 1.2:1) and raise the release to about 250 ms to start. From then on, it's a matter of fine-tuning attack, release, and ratio, and possibly readjusting the threshold. The object is to put the threshold in between the lower and higher dynamics, creating a constant alternation between high and low (or no) compression within the music. Don't forget that the gain reduction meter generally lies: 1 dB of metered gain reduction can mean a lot. Note that too-low a threshold will defeat the purpose, which is to differentiate the "syllables" of the music: with too low a threshold, *everything will be brought up to a constant level.*

It's unusual to see such low ratios in tracking and mixing, but very common in mastering, partly because with full program material, larger ratios may create audible breathing, pumping or other artifacts. Typical thresholds are in the -20 to -10 dBFS range. But there is no rule: some engineers get great results with ratios of 5:1, whereas a delicate *painting* might require a ratio as small as 1.01:1 or a threshold of -3 dBFS. One trick to compress as inaudibly as possible is to use an extremely light ratio, say 1.01 to 1.1, and a very low threshold, perhaps as low as -30 or -40 dBFS, starting well below where the action is. In this case the compressor is not bouncing on the syllables but rather giving a gentle, continuous form of macrodynamic reduction. We may choose a low ratio to lightly control a recording that's too *jumpy*, or to give a recording some needed *body*. Another trick to get invisible compression is to use some parallel compression, as described in the next chapter. With limiters, I try to never let the limiter do the heavy lifting — I rarely use more than 1 dB of limiting — and rely on the combination of processors in the chain to accomplish the lifting.

Going Beyond?

Every recording has a point that I call its "loudness potential," above which we can only hurt the sound by adding too much distortion, or losing the impact and the beat. This point of degradation is quite obvious when auditioning on a good monitor system, where we find that even 1 dB more level will clearly hurt the

"A Vari-Mu that can produce some whip along with the cream!"

sound. But when a client insists I turn it up even when I tell him the recording has already reached its loudness potential, we have to turn it up or lose the job — even though it will be less "radio-ready" (see sidebar page 95). In the case of today's most distorted genres, or when the client insists that the CD make his whole car jump, it's pretty hard not to break every single rule. The loudness issue has come to the point where if the client wants it "stupid loud" (my term, not theirs!), I may have to run a number of items in my chain into distortion to satisfy him — and I cannot apply any degree of sophistication or refinement to satisfy his needs. In Chapter 9 I discuss some subtle techniques to make things a bit louder without hurting the sound too much, but once we reach "stupid loud," even those techniques no longer apply. In Chapter 17, I discuss how and why recordings in the immediate future will sound much better!

Compressors With Unique Characteristics

Part of the fun in using compressors is discovering the specialities of different brands and models. Even with the same settings, some are *smooth*, some *punchy*, some nicely fatten the sound, and others make it brighter, harder or more percussive. This is often due to differences in the curve or acceleration of the time constants (attack and release times), how the device recovers from gain reduction, and whether the gain returns to unity on a linear, logarithmic, or even an irregular curve.

Analog compressor designers choose from several styles of gain manipulation. The most common are **FET** (field effect transistor), **optical** (abbreviated **opto**), **VCA** (voltage controlled amplifier), **Vari-Mu**, **PWM** (pulse width modulation) and their various subcategories. Digital designers may emulate their characteristics, as in the Waves Renaissance series of digital compressors that have both *opto* and *electro* modes. As mentioned before, in opto, the release time slows down for the last portion of the release, while in electro it accelerates. Electro can yield a more aggressive sound, while opto is good for gentle, easy-going purposes. Analog optical compressors are great on vocals in tracking or mixing, but not as good for aggressive mastering of overall program material because they are generally too slow. However, digital opto models can be faster than their analog counterparts.

Generally, analog optical models are more suitable for "gentle" mastering. However, one model, the Pendulum OCL-2, has a proprietary optical sensor whose reaction time is much faster than others on the market. It also has a very transparent tube circuit that can provide very subtle warming. This makes the OCL-2 perhaps the only optical analog compressor with the gentleness of optical (useful for adding body), but can be set fast enough to provide a bit of punch. However, it is not as fast as a VCA- or PWM-based compressor, so it may not be capable of achieving enough "attitude" or punch in aggressive music.

In my opinion, closest to a Swiss Army Knife is the Crane Song Trakker, a solid-state compressor that can emulate the tonality and speed characteristics of several different types of compressors and embody some of the warming characteristics associated with tubes. The only downside of the Trakker is the learning curve (the best way to learn is to experiment with each of the presets).

Another compressor feature is to add supplementary low frequency harmonics to change tonality, as in Waves "warm" setting of the Renaissance Compressor plug-in. Note that the addition of harmonics slightly compresses sound by reducing the peak-to-average ratio. Waves calls the alternative setting (when the extra harmonics are turned off), "smooth", but I think that's misleading, since the term "smooth" can be confused with the attack/release characteristics.

Another style of compressor is the Vari-Mu, whose ratio varies with level, a circuit popularized by Manley. Vari-Mu is especially helpful for creating lush and "creamy" atmospheres in ballads and popular music in the Broadway and pop-classical fields. Vari-Mu's particular dynamic characteristic preserves micro-dynamics and transients, while manipulating the macrodynamics, unobtrusively bringing up low passages. Typical Vari-Mu action may be slow for enhancing music with fairly fast dynamics, but the Pendulum ES-8 is a unique Vari-Mu that can be made slightly faster to produce a bit of whip along with its cream!

Sidechains

Most of the time the sidechain (control path) is identical to the audio signal, but interesting things can happen when it is not. For example, in a stereo or multichannel compressor, each channel has its own sidechain, but it is possible to feed or *link* all sidechains from one channel's signal. By linking the sidechains, one channel controls the gain reduction of both equally. Alternatively, the design can sum the channels into the sidechain, so that the higher of the two brings down the level. In my experience, the latter technique unfortunately reduces stereo separation. The linking switch prevents image wandering. Without it, if a drum hits much louder in one channel than the other, the image will momentarily move towards the opposite channel. When unlinked, the box operates as two independent mono compressors. In some models, unlinked is labeled *dual* and linked is labeled *stereo*. In multichannel compressors, there may be a separate sidechain for front and surround, or all channels may be linkable.

The textbooks tell designers to link the compressor channels, and while that seems desirable, I've found that apparent stereo separation can increase if the linking is removed. It's hard to say whether this is a psychoacoustic improvement or a technical improvement, or both! Don't be afraid to use the independent sidechain mode for stereo if it sounds better to you. When unlinking, just be careful to check for image wandering, which may even be desirable and add an artificial sense of space. With the small amounts of compression I usually use, I've never noticed an image shifting. I've also found that running digital limiters unlinked can reduce clamping effects, i.e., where the sound appears to drop and not recover fast enough (because the channel that did not drop will mask short-duration drops in the other). Therefore, unlinking limiters helps make things sound both louder and cleaner. As with the unlinked compressor, watch out for instantaneous image shifts caused by extreme transients located only in one channel. But you would never use more than 1 dB of limiting yourself, right?

Sidechain EQ: Often, sidechains are fed an equalized signal. Perhaps the most popular sidechain EQ is a high-pass-filtered signal, which helps prevent the bass drum from pushing down (or *modulating*) the rest of the music. In an analog compressor, it is very easy to implement this without needing a separate equalizer, simply by inserting a capacitor (approximate value 0.1 µf) in series with the sidechain inserts: the exact value depends on the source impedance. Measure the amount

The Real Recipe for Radio-Ready

The real recipe for Radio-Ready includes:

1) Write a great original song, use fabulous singers and wonderful arrangements.

2) Be innovative, not imitative.

3) Make sure the music sounds good at home. Keep the dynamics lively, interesting and unsquashed, and some of that virtue will make it through the radio processing.

"Fix the disease at its source, not with a multi-band-aid in mastering."

of gain reduction with a test tone and experiment. It's useful to have a switch to add or remove the filter. In the Pendulum units, side chain is on a TRS plug, so a simple capacitor wired between tip and ring of a plug creates an instant sidechain filter. I've been conducting a friendly competition with a fellow mastering engineer who claims that the most natural-sounding compressor sidechain should follow the ear's loudness-sensitivity curve, while I maintain that a simple low cut is sufficient to get bass drum punch.

With a sidechain boost in the upper mid or low treble range, the compressor becomes a **de-esser**, or it can be tuned to deal with troublesome cymbals. The only problem with sidechain-based de-essing is that the entire range of audio frequencies is brought down whenever an "s" goes by, so generally the gain reduction has to be kept subtle, no more than a dB or two, unless you are looking for a special effect. Dedicated de-essing requires very different time constants from mastering for punch; generally with an s you want to be surgical, attack it quickly, and recover just as quickly.

IV. Multiband Processing

Multiband: Advantages and Disadvantages

Multiband processing was probably first introduced by TC Electronic, who was also the first company to introduce linear-phase band-split filters (linear phase is the only way to do multiband compression correctly, unless the unit uses first-order crossovers). They started with their M5000, then the ubiquitous Finalizer, and brought multiband to great sophistication, subtlety and versatility in the System 6000 with the MD4. But for most downward compression purposes multiple bands are rarely needed: one or two bands are usually enough. The Weiss DS1-Mk3 has one active band; the compressed signal can be isolated to one frequency range as surgically as necessary and the rest of the spectrum left unaffected. Until I got the Maselec MLA-4, most of the compression I performed in mastering was with a full-band compressor, or a full-band compressor with a high-pass sidechain, or the Weiss with the active band usually above some low frequency. Rarely does any recording need more than one active frequency band to sound punchy and strong.

However, splitting a compressor's signal into multiple bands (and multiple sidechains) avoids the problem of modulation with a single sidechain, since compression in one band will not affect another band. For example, the vocal will not pull down the bass drum (or vice versa). This is perhaps the biggest selling point of multiband, because with the same amount of gain reduction it can sound superior to wideband or sidechain equalization. The action can be made very invisible. Also, a higher amount of compression and average level can be achieved in a multiband with fewer interaction or "clamping" artifacts. Another advantage is that high frequency transients can be left unaffected while compressing the midrange more strongly, producing a brighter, snappier sound than a single band unit. But when the thresholds are set aggressively, loud action in one frequency band can dynamically change the overall tonality, producing a noncohesive sound — especially if all the bands are moving in different amounts throughout the song. In the analog Maselec MLA-4, the sound is sweeter than most digital multiband units, partly because of its first-order (6 dB/octave) gentle slopes. It's easier to be subtle and gentle with this unit. It's also possible to link the band sidechains and make it perform like a single band unit.

Multiband units make good sound "sweeteners". When compression is applied solely in the high frequency band, the sound gets duller as it gets louder. You can use this technique to:

- Construct an analog tape simulator
- Sweeten the harshness of poorly-recorded digital recordings
- Soften distortion that gets harsh when it gets loud
- Sweeten bright cymbals without softening the average tonality (as opposed to an equalizer)

In the case of bright cymbals, we can take advantage of the fact that they are usually located on the right or left side rather than the center, and compress just the high frequencies of the side channel to sweeten the sound. This technique can be very surgical, effective, and unobtrusive (see Chapter 9).

Multiband units make good de-essers. Sibilance can be controlled by using selective compression in the 2 kHz through 10 kHz range. The actual frequency has to be tuned by listening to the vocalist. Some de-essers allow monitoring the sidechain to help find the offending frequency: try starting at 3 kHz. Try a very fast attack, fast release, **crest factor** set to peak (to be explained), and a narrow bandwidth for the active band. The Weiss DS1-Mk3 is hands-down the best-sounding mastering de-esser I've encountered probably due to its peak sensitive compression, narrow linear phase band-split filters, dual-speed release, M-S or L-R option (new for the Mk3), and very effective presets. Mastering de-essing is the acid test for a processor: most devices I've tried are either not invisible because they aren't selective to just the s sounds, or on the other hand, they miss too many s's.

The multiband device's virtues permit louder average levels than were previously achievable – making it the most powerful but also potentially the most deadly audio process ever invented. "Deadly" because multiband compression fuels the loudness race (see Chapter 17). The technique has been hyped as a cure for all ills (which it is not), and it can easily produce a very unmusical sound or take a mix where it doesn't want to be.

The key to a great master is a great mix!

Generally I do not try to "remix" with multiband, simply polish what is there. However, multiband can be used to help improve ("repair") a bad mix when a remix is not possible, and some mastering engineers become experts at this technique. I once received a rap project that was mixed with very low vocal, and extremely loud percussion and bass drum. A remix was not possible, but by compressing and then raising the level of the frequencies in the vocal range (circa 250 Hz), I was able to *rebalance the piece* and turn the vocal up. Just don't be fooled into believing that toning down loud instruments through multiband compression is the same as a remix. Using a band-based compressor squashes the instrument's or vocal's sound as well as allowing you to manipulate the level within its frequency range. However, with delicate technique I have invisibly helped a big band recording that was a good mix, except in loud passages the singer's level was sitting a bit too loud. During the loud passages I cheated the singer down only 1 dB with a single M-channel compressor in the midband, and brought the instruments up with upward expansion on the S channel at the same time. This turned an A mix into an A+ master, with more impact for the band and without losing any vocal clarity. The next Chapter in our dynamics trilogy will explain upward expansion. Chapter 9 will detail M-S processing.

Before trying to "remix" with multiband, try these tips:

- Add a little bit of equalization in the frequency range of the instrument that needs to be raised. Make sure the "cure" sounds uncolored and better than the "disease."
- See if simply raising the attack time in a one-band compressor permits sufficient transient energy to come through. Or, try upward expansion instead (See Chapter 7).
- Use fewer bands – only two if possible, to reduce artifacts.

Equalization or Multiband Compression?

When multiband processing is used, the line between equalization and dynamics processing becomes nebulous, because the output levels of each band form a basic equalizer, and if the bands have different thresholds, then they affect the tonality dynamically. Use standard equalization when instruments at all levels need alteration, or use multiband compression to provide spectral balancing at different levels. This is a form of dynamic equalization, so depending on one's point of view, a multiband compressor can be looked upon as a dynamic equalizer.

When already using a multiband unit, we can make our first pass at equalization with the outputs (makeup gains) of each band. Multiband compression and equalization work hand-in-hand. Tonal balance will be affected by the crossover frequencies, the amount of compression, and the makeup gain of each band. With few exceptions, the more compression, the duller the sound, so first try to solve this problem by using less compression, or by altering the attack time of the high-frequency compressor. As a last resort, use the high frequency band's makeup gain or an EQ to restore the high-frequency balance.

"Never in the history of mankind have humans listened to such compressed music as we listen to now."
— Bob Ludwig [2]

In summary: Use multiband for remixing as a last resort. Fix the disease at its source, not with a multi-"band-aid" in mastering! As a matter of course, most of the tools we can apply in mastering to try to remix often sound worse or compromised compared to a remix.

V. Refinements

Alternatives to Downward Compression

High frequency equalization post-compressor can be used to alter the perceived level of snap (transients). Upward expansion used alone or in conjunction with downward compression (see Chapter 7) can help restore or enhance some of the snap. But there is no cure for **hypercompression**, only a band-aid.

Emulation vs. Convolution

Digital compressor designers have the choice of **emulating** the transfer characteristics of a source compressor and implementing them in DSP, or sampling the source compressor and **convolving** that sampled characteristic with the incoming audio. Convolution works well with reverbs and equalizers, but I have not personally encountered a successful convolution-based digital compressor. These are very difficult to implement: the source device has to be sampled at many different levels to ascertain its dynamic characteristics, which are then very hard to interpolate accurately and with the right timing. The convolution processor has to be fast, and operate with great resolution in the oversampled domain to prevent **digititis** (edginess in the sound caused by artifacts of digital processing).

Fancy Compressor Controls

Some compressors provide a **crest factor** control, expressed in decibels, or a range from RMS (or full average) to quasi-peak through to full peak. This means that the compressor can be set to act on the average parts of the music, the peak parts, or somewhere in between. Ostensibly, compressors with RMS characteristics sound more natural, because they correspond with the ear's sense of loudness — but one of the best-sounding compressors I know is peak-sensing. When the crest factor control is set to peak, short transients tend to control the action, and at RMS, more continuous sounds control it. The TC MD4 has a continuously-variable crest factor control. In most cases leave it at RMS, but for de-essing and for better control of transients (such as a too-loud snare drum or softening cymbal hits), move it closer to peak.

The Weiss model DS1 has two different release time constants, called *release fast* and *release slow*. The user sets a threshold of average transient duration, such as 80 ms, above which a sound movement is called *slow*, and below which it is called *fast*. Instantaneous transients receive a faster release time, but sustained sounds receive a slower one, which results in a more natural-sounding, yet louder compression. Indicator lights on the front panel make adjusting this a snap.

A few processors have multiple thresholds available. I currently own one, originally conceived by Roger Nichols, then passed to RN Digital as the models D1 through D4. I've successfully used the D4 in a case where nothing else seemed to work, downward expanding the low area as a form of noise gate to soften some lighting buzz, upward compressing the midpart of the dynamic range to bring out inner details, and upward expanding the top part of the range to restore some impact of the loud peaks, all in one plug-in. The interface of the D4 is very intuitive, allowing gradual overlap between the different areas. Though I may use it only once in a year, it's been a life-saver. Other ways of using multiple thresholds would be to "overlap" two compressors in series: for example, the first compressor performs a gentle overall compression with a low threshold and ratio, and the second more aggressively controls some offensive peaks at high levels.

"Not every instrument should be up front."

In a related fashion, it's not uncommon to run several dynamics processors in series in mastering, each doing a small part of the job. To keep a peak limiter from clamping down on the signal too much, try preceding it with some form of *preconditioner*, which may be an analog compressor that can gently soften a transient so that the limiter which follows doesn't have to work as hard.

Clipping, Soft Clipping and Oversampled Clipping

Digital Clipping is the result of attempting to raise the level higher than 0 dBFS, producing a square wave, a severe form of distortion, also known as "clipping the medium". *Analog clipping* is the result of overdriving an analog processor beyond its peak headroom. Analog clipping generally comes on gradually, while digital clipping becomes severe quite rapidly. Therefore, analog clipping sounds easier on the ear than digital clipping, but they are both forms of distortion. **Clippers** are specialized devices that electronically cut momentary peaks out of the waveform to allow the overall level to be raised, ostensibly a better-sounding approach than just clipping the medium. *Soft clipping* attempts to do this only a little bit, with less distortion, but I think the results have been underwhelming. As you can guess, I'm not a big fan of digital clippers, even the ones that claim to get rid of **aliasing** artifacts by **oversampling** their processing (aliases are ugly-sounding beat notes between the sampling frequency and the harmonics of the distortion). If I do have to clip to get more level, I prefer to clip an analog stage: the harmonic distortion products produced are more organic-sounding to the ear than any digital clipper I've encountered, and the filters in the ADC get rid of harmonics above the **Nyquist frequency** without causing aliasing. According to Jim Johnston, it would take far more than 8x oversampling to purposely clip digitally without causing audible alias artifacts, which most designers don't do, and many CPUs can't handle (yet). The better approach is not to raise the level at all, for many CDs are already too hot for their own good. Clipping is also a severe problem for codecs such as AAC and mp3, which simply don't like it. So if you still want to clip, first consider the effects when coding. In Chapter 16, we'll look more closely at types of clipping, why it should not be acceptable, but why sometimes we can get away with it. In the meantime, I do not advise mixing engineers to clip ADCs to get more "attitude" — at least until you have an objective shootout with a mastering engineer, sending both clipped and unclipped mix versions to see which one sounds better after the mastering stage. Remember that distortion is cumulative, and it doesn't take much to send a good-sounding recording over the edge.

Compression, Stereo Image, and Depth

Compressors tend to amplify the mono information in a recording. Unfortunately, compression also collapses the soundstage of a recording while it brings up the inner voices in musical material. Instruments that were in the back of the ensemble are brought forward, and the ambience, depth, width, and space are degraded. Not every instrument should be "up front." Pay attention to these effects comparing processed vs. unprocessed and listen for a long enough time to absorb the subtle differences. Variety is the spice of life.

The Mastering Engineer's Dilemma

Without compressors in playback media, it is extremely difficult for the mastering engineer to fulfill the needs of both casual and critical listeners, whether sitting in quiet living rooms, jogging, or driving in a noisy car. It is our duty to satisfy the producer and the needs of the listeners, so we should continue to use the amount of compression necessary to make a recording sound good at home. But try to avoid using more compression than is required for home listening; this will actually help FM radio play by reducing distortion artifacts in the radio processors. If compromises have to be made for car or casual play, try transparent-sounding techniques for raising soft passages, such as parallel compression (See Chapter 7).

In Conclusion

Learning the nitty-gritty of compressors is a lot of fun and will be a lifetime task. The more I use them, the more I learn.

1
One manufacturer, DBX, measures release time in dB/second, which is probably more accurate, but hard to get used to.

2
In correspondence. A variation of this quote is in Owsinski, Bobby (2000). *Mastering Engineer's Handbook.*

CHAPTER 7

How To Manipulate Dynamic Range for Fun and Profit: Think Forward

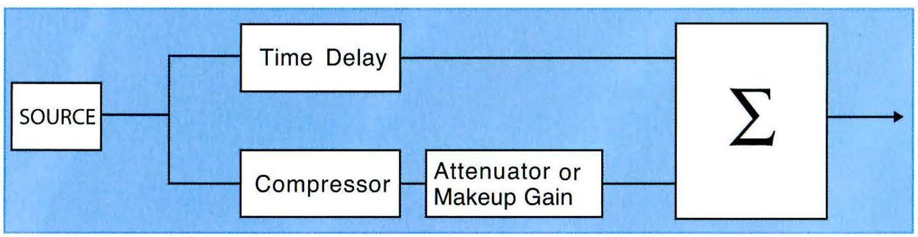
Introduction

In this Chapter, we'll introduce two new types of dynamics processors that seem counterintuitive: to make the second one work, you have to learn to think "forward" instead of "backward." The effort is well worth it.

I. Upward Compression

When compressing a recording, I believe the ear is much more forgiving of "cheating" soft passages upward than of "pushing down" loud passages. The latter, downward compression, can sometimes diminish the punch, snap, impact and clarity of a recording, while the former, upward compression, can feel more natural. As usual, the devil is in the details.

Let's introduce an upward compression technique that requires just a single knob — no need to adjust attack, threshold, release or ratio – and it's so sonically transparent that only careful listening reveals that the circuit is in operation! New Zealand radio engineer Richard Hulse described his practice of **parallel compression** to me: He had been using analog components and was getting acceptable results, but suggested I try a digital implementation.[1] The digital approach proved so successful that I use this mastering technique frequently – either to supplement other techniques in more aggressive music styles like hip hop, or as the sole compressor in more subtle styles like classical or jazz music.



The Parallel Compression technique employs a matched time delay in the "dry" signal path to avoid phase shift or comb filtering. Σ is the symbol for "sum." This yields very transparent-sounding upward compression.

The principle is simple: take a source, and mix the output of a compressor with it. In the digital domain, it is possible to sum the source with a compressor without any side effects, by using a precise time delay for the "dry" signal exactly matching that of the compressor, as shown in this block diagram (below, one channel only of stereo shown).[2]

The parallel compression technique produces less distortion than standard (downward) compression, because the path is divided in two parts. For example, if the compressor and the dry signal are mixed equally, the distortion is reduced by 6 dB (one half), compared to using the compressor alone.[3] The effective amount of compression is controlled either by the attenuator that mixes the compressor with the linear path, or by the compressor's own makeup gain. When you build a parallel compressor from plug-ins, a DAW with automatic latency compensation does not require the extra time delay. Test this by adjusting the parallel compressor to a 1:1 ratio and unity gain, and invert the **polarity** of either half of the chain. This will produce a complete null (no sound) if the time delay is correct to the sample. Or you can simply move the compressor gain up and down to verify that there is no comb filtering.

Transparent Parallel Compression

I have two different approaches to parallel compression. The first one is the *transparent* approach taken by Richard Hulse, in which the compressor is as invisible as possible, producing no obvious tonal shifts and little or no loss of transients. In many cases this technique is indistinguishable from manual fader riding, and highly suitable for delicate acoustic music. This transparent parallel compressor raises gain at very low levels and contributes less to the total sound as the signal gets louder. Here are the ingredients of this recipe:

- *Threshold*: set the parallel compressor's threshold to an extremely low -50 dBFS. This will put the parallel compressor into heavy gain reduction nearly all the time, and ensures that the compressor will be applying extreme gain reduction during loud passages. This works because, in principle, if you add a second signal that is 20 dB or more below the main signal, this second element will not perceptibly contribute to the total level or the sound. Because the output of the parallel compressor is pushed significantly down during loud passages, it contributes only negligibly at high levels.
- *Attack time*: set as fast as possible – one millisecond or less, if available. This will preserve the transient impact of the original sound: as soon as a loud transient hits, the compressor will instantly go into gain reduction. The faster the attack, the more invisible the parallel compressor, and more transients are preserved, requiring accurate look-ahead (see Chapter 6). Note that this is opposite to how you would use a downward compressor, where very short attack time softens transients. Here we must think "backward."
- *Ratio*: set to 2:1 or 2.5:1 (I prefer 2.5). This is the root setting of the compressor, but actual achieved ratio in parallel depends on the output level of the parallel compressor. Richard has developed a chart by the numbers, but I prefer to go by ear.
- *Release time*: set to medium-length. Experiments show that 250-350 milliseconds works best to avoid breathing or pumping, although in cases where the reverberation is very exposed, particularly in *a capella* music, as much as 500 ms may be needed to avoid overemphasizing the reverb tails.
- *Crest factor*: set to Peak. I've found the most transparent parallel compressors are peak-sensing.
- *Output level* or *makeup gain*: adjust to taste. With the parallel compressor off (-∞ gain), there is no compression. Above about -5 dB, compression will be very noticeable, with even medium-level passages being raised in level, and there will be an audible loss of transients and snap. A nice subtle compression can be achieved with settings of -15 through -5 dB.

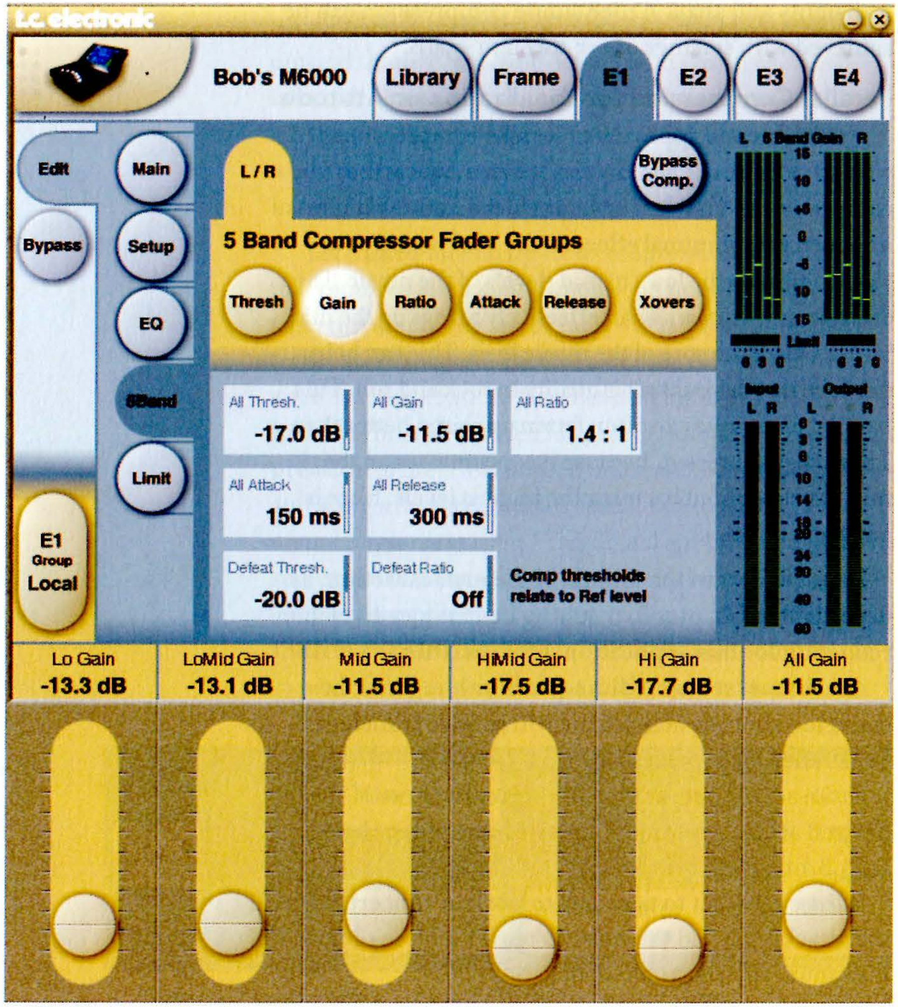
Parallel Compression for Tonalization or Attitude

In this second approach to parallel compression, which I call *attitude parallel compression*, we set the compressor in a normal way, to achieve some attitude or punch with minimal effect on the loud peaks, or to warm or clarify the low- to mid-levels of the music. The *attitude parallel compressor* effectively brings up the mid levels, where the core of the music lives. This can help produce that desirable *epic* quality in rock and roll. This parallel technique can often fatten up sound better than a normal compressor, because it concentrates on the mid-levels without harming the highest levels. Here is our second recipe:

- *Threshold*: set in the middle of the musical action, as described in Chapter 6, resulting in *up to* 5-7 dB of gain reduction for this application, often as little as 1-3 dB.
- *Attack time*: set to medium (start with 125 ms), because too short an attack time will subdue transients. Look-ahead is unnecessary.
- *Ratio*: set to taste, set for the aggressiveness of the desired action, in conjunction with the output level. Try 1.3:1 through 3:1.
- *Release time*: set to taste, set to work in concert with the attack time and the music's movement, to obtain maximum rhythm and punch.

• *Crest factor*: set to RMS. I've found the best attitude compressors are RMS-sensing.

• *Output level* or *makeup gain*: adjust to taste, whatever obtains the desired effect, rarely past -6 dB. Even mixed at -6 dB, parallel compression will degrade transients and instantaneous peaks less than a standard downward compressor, because the dry signal is still contributing more to the total than the compressed signal.



Tonalization: The TC MD4 performing "warming" parallel compression using higher gains in the low mid band.

Tonalization is my term for a form of dynamic equalization performed by using a multiband compressor in parallel mode. The tonality of the program can be changed by manipulating the gain of each frequency band being mixed in parallel. For example, you can warm up the mid-levels of the program without sacrificing clarity at high levels, or add presence at low levels. The other advantage to parallel bass-frequency compression is that the body of the bass instrument gets fatter without destroying its pluck. Or when you increase the presence frequencies at low levels, the sound can be clearer and better defined without becoming harsh at mid or loud levels.

Several digital compressors incorporate built-in parallel compression, including the TC Electronic MD4 in the System 6000, Weiss DS1-Mk3, and the PSP plug-in Mastercomp. The Weiss's action is imperceptible, so it's ideal for classical music. The MD4 excels at both tonalization and attitude style and is no slouch at the transparent style. For subtle fattening, which also contributes to the punch, the MD4 is my "go to" multiband parallel compressor: typically having 3 to 5 dB of action and mixed as low as -14 to -17 dB. Multiband simply helps to keep it more invisible than single band.

The MD4 (pictured here) can also *tonalize* by subtle manipulation of the levels of each band, via the faders at the bottom. Additional warming is accomplished via a little more parallel compression in the low and mid bands than in the high bands. These are the gains at which the output of the compressor is being mixed with the dry signal. This produces a very different effect than a normal downward multiband unit because it subtly brings up low level material in selective bands but leaves the high level material relatively untouched. However, if you hear a radical tonal difference at different levels, set the band gains closer to the same amounts.

As with any process, if upward compression is pushed too far, it will call attention to itself. It's like using fill flash on a camera: too much fill, and the picture becomes overexposed. The first audible artifact will be increased sustains and emphasized reverberation, followed by loss of transients, and finally, breathing or pumping. Instant hip hop hit! These artifacts can sometimes be reduced by raising the release time or raising the attack time when using the attitude approach. However, if the music is so open or delicate that the process continues to call attention to itself, the only solution is to abandon the processor and manually raise the passages which are too soft. Also note that not every song that needs compression benefits from maintaining transients. Rock and roll music can get punch and power from judicious downward compression with medium or long attack time.

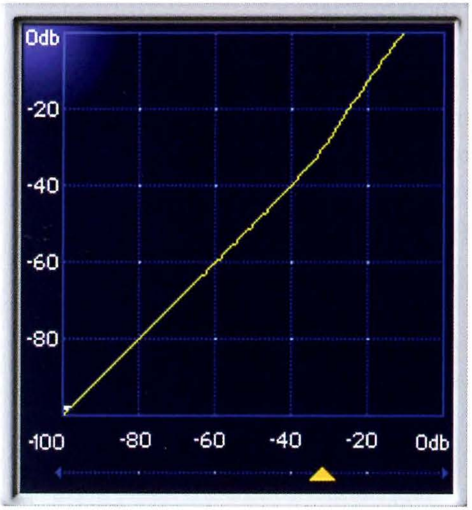
II. Upward Expansion

Another underused but incredibly useful processing technique is **upward expansion** – definitely a technique worth learning, and no more difficult to use than a downward compressor, once you learn to *think forward*! Some people think of upward expanders as **uncompressors**, but they're far more than that (indeed there is a limit to how much a sound can be restored once it has been excessively compressed). Rather, upward expanders can be used to emphasize different parts of the dynamic rhythm than those parts affected by downward compressors. For example, upward expansion is great for adding liveliness to boring musical samples, and it can also put the snap back into a slightly-squashed snare drum. I personally avoid devices like transient enhancers or exciters, which add their own distortion. Instead, when searching for more snap, I use an upward expander to directly manipulate or enhance the microdynamics.

In the analog days, upward expanders were difficult to build until the advent of the VCA,[4] but it is a simple matter to turn any VCA-based compressor into an upward expander by inverting the control voltage. The first commercial dedicated upward expander was probably that built into the DBX model 117 (1971), which was designed to enhance dynamics in a hi-fi system. Another early upward expander was the Phase Linear Peak Unlimiter. The honor for the first digital upward expander goes to the Waves C1 (plug-in), algorithms designed by Michael Gerzon; ever since then, every Waves dynamics processor includes fractional ratios for upward expansion. The first stand-alone digital upward expander was in the DBX Quantum mastering unit, followed shortly by the Weiss DS1-MK2. The Waves C4 (plug-in) is the first single processor that can perform all four dynamics processes. The Maselec MLA-4 (page 108) is the first analog multiband processor which can perform simultaneous downward compression and upward expansion in different bands, very useful, e.g. to de-ess the high end and at the same time add attack and punch to the bass drum.

Think Forward

In downward compression the output level is pushed *downward* while the incoming level is moving *upward* (above the threshold), which is out of sync with the natural movement of the music: applying too much downward compression is like trying to swim away from

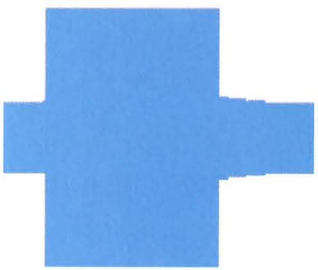


An upward expander with .75:1 ratio, expressed in decimal (1:1.33 expressed as a fraction). Threshold is -32 dBFS, and without attenuation, the output will overload if input exceeds approximately -10 dBFS.

shore as heavy waves push you backwards towards shore. By contrast, upward expansion *further increases* the level of incoming passages that are already *increasing* – a process that *is* in sync with the motion of the music. It's like swimming toward the shore while being pushed *forward* by the waves. Though it may be necessary to use output attenuation instead of makeup gain to prevent the output from overloading. Upward expansion results in an *increase* in dynamic range: if used appropriately and delicately, it becomes as valuable a production tool as downward compression. Neither process is the cure for every problem, but today in mastering I often have to think forward instead of backward!

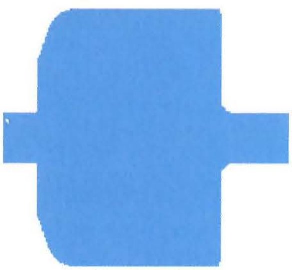
For illustration (pictured above left), this transfer function shows an upward expander with a severe .75:1 ratio and threshold at -32 dBFS. Without attenuation, it will overload with input levels exceeding about -10 dBFS. Note that the ratio of an upward expander can be expressed in decimal or fraction form, depending on the manufacturer's preference. The Waves units use decimal form, while the Weiss unit expresses this ratio as a fraction, 1:1.33. Typically, the usable range of ratios for mastering with upward expansion is small, from a very gentle 1:1.01 through about 1:1.2 (fraction); equivalent to 0.99 through .83 (decimal). A useful value for music enhancement is around .95 decimal (1:1.05 fraction).

Figure A



Upward expander with fast attack, slow release.

Figure B



Slow attack, fast release.

Figure A (below) shows an upward expander with fast attack and slow release; Figure B shows one with slow attack and fast release. As you can see, the dynamic characteristics are the opposite of the compressor examples shown in the previous chapter. The best way to learn how to use an upward expander is to compare it to a downward compressor, described in the chart on page 109.

Upward compressors and upward expanders make a nice team that can fatten sound at low levels, raise the average level, and increase the impact or liveliness at high levels. You can also combine downward compression with upward expansion using the right combination of thresholds, thus relegating the compressor to bringing up the mid-levels, and the expander to produce more dynamic and natural high levels and preserve the transients. While downward compressors tend to "sweeten" or warm up sound, upward expanders tend to brighten sound, because they increase the strength of transients. If the sound becomes too bright or "snappy," consider expanding only the bass through the lower midrange. This warms up the sound *by raising the bass as loudness increases*. This is subtly different from a tape saturator, which *lowers high frequencies as loudness increases*. The former approach enhances dynamic impact, while the latter may soften it.

Compromises When Making Hot Masters

Neither downward compression nor upward expansion works very well above a certain program loudness.

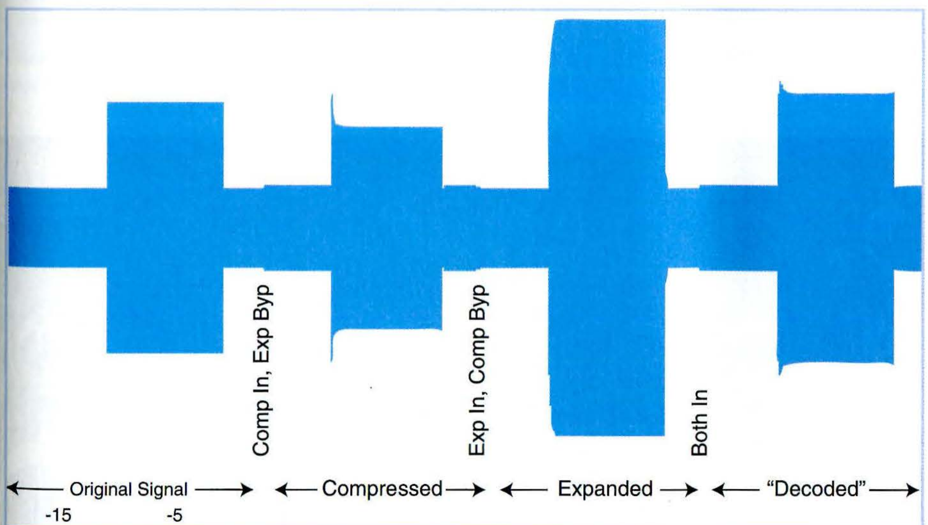
With the first technique, when we are asked to make a "louder" master, the sound becomes more squashed. With the second technique, to prevent peak overload distortion, we have to use more limiting. This eventually counteracts the expansion, and impact is again lost. If we cannot live with the degradation, the only solution is to master at a lower program level.

Measuring the Effect of the Processing

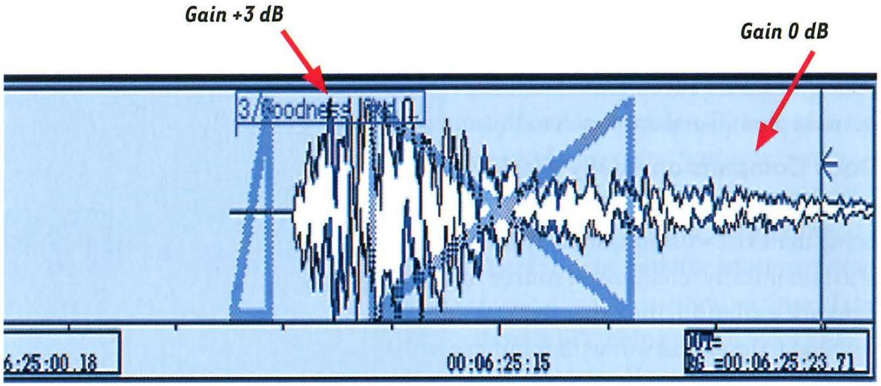
A few level meters have been developed that simultaneously show the average and the peak levels of the music. Pro Tools 11 includes the K-System meters, which can do this. If the distance between the peak and the average has increased, chances are that the microdynamics have also increased. During mixing, simply raising the snare drum level a dB can increase the peak-to-average ratio of the mix. During mastering, raising the attack time on a downward compressor or lowering the attack time on an upward expander can raise a snare drum's apparent level as well as the microdynamics of the entire piece. In Chapter 16 we will discuss a new concept called **PLR** (peak to loudness ratio), which is a more sophisticated approach to this measurement.

Does Companion Really Work?

Is there such a thing as an "uncompressor" that can bring material with squashed dynamics back to life? If a dynamically-challenged source has no incoming dynamic variation, the expander can do nothing, or it will make the sound worse. But if there is some dynamic movement left in the source, an expander with the right parameters can improve its movement, pace, rhythm, and transient impact. This requires experience and careful listening. **Companding** means *compression followed by complementary expansion*. The figure (below) shows companding can work with some loss of information. A toneburst alternates between -15 and -5 dBFS, followed by a downward compressor, then a complementary upward expander. As illustrated, the average levels are restored, but the initial transient attacks, and to some extent the decays are not well-preserved.



Does Companion Really Work?



Creating an artificial sforzando

III. Increasing Microdynamics Manually

We can change musical macrodynamics, and sometimes the microdynamics, by doing manual edits and gain changes in a DAW. In the above figure, the attack of the first note of a song has been artificially enhanced with very brief manual upward expansion (the brevity makes it microdynamic). At left, the first few milliseconds of the note have a greater gain (in this case, 3 dB), and then there is a crossfade to a gain of 0 dB, resulting in a **sforzando**. Interestingly, the producer was looking for a surprise when this track entered, and I initially had the beginning attack at +5 dB. But when he took the reference CD home, he was really startled, so I took it back a bit for the final master.

This Chapter completes our dynamics trilogy.

1
Richard initially called this *sidechain compression*, but I suggested a name change to avoid confusion with the sidechains of compressors. This technique was publicized by Mike Bevelle in the article Compressors and Limiters, **Studio Sound**, October 1977 (also reprinted June 1988). Also known as "New York style compression," engineers have been playing with parallel compression techniques for many years.

2
There are other theoretical methods for achieving upward compression: for example, summing a downward expander with the dry signal in opposite polarity; or upward expansion, summing a downward compressor with the dry signal in opposite polarity. I have not experimented much with either technique. Thanks to Cris Allinson for this hint.

3
This was the principle of the Dolby A/SR systems, which used a direct signal path summed with a compressed one, *doing as little harm to the audio as possible*.

4
Voltage controlled amplifier. When automating a mix in an analog console, the audio passes through a VCA unless it has moving fader automation.

The Maselec MLA-4 is the first analog multiband mastering processor with each band switchable between downward compression or upward expansion.

DOWNWARD COMPRESSION	UPWARD EXPANSION
makes sound louder during the **descent** of the music (release phase)	makes sound louder during the **rise** of the music (attack phase)
makes the mid-levels of the music louder and the high levels softer	makes the high-levels of the music louder and the low to mid levels softer
tends to make sound fatter and exaggerate low frequencies (subject to time constants and threshold).	tends to exaggerate transients and high frequencies (subject to time constants and threshold).
Attack times that are too short (fast) cause transients to be lost.	Attack times as short as a few ms can restore and sharpen lost transients (e.g. from analog tape or overcompressed sources).
Typical attacks 100 ms through 300 ms. Less than 40 tends to soften or blur transients.	Typical attacks 1 ms through 300 ms. If a transient still sounds too sharp with >150 ms attack, perhaps this is not the right process for this music, or consider a touch of limiting after the expansion, or expand only below a certain frequency.
tends to make things sound duller or warmer.	tends to make sounds brighter or sharper.
tends to go **against** the natural movement of the music, especially when the parameters are not optimized.	tends to work **with** the natural movement of the music, especially when the parameters have been optimized.
If sounds "jump out" too much, raise the ratio, shorten the attack, and/or speed up the release.	If sounds "jump out" too much, lower the ratio, lengthen the attack, and/or slow down the release.
If attacks seem too sharp, shorten the attack time.	If attacks seem too sharp, lengthen the attack time, use less expansion, or apply expansion only below a certain frequency.
If sustains seem too long or too prominent, lengthen the release time.	If sustains seem too short, lengthen the release time.
If attacks seem too dull, lengthen the attack time.	If attacks need enhancement, shorten the attack time.
If you don't like the percussiveness (e.g. snare drum), speed up the attack. To increase the ratio of rhythmic snap to smoothness, lengthen the attack. Downward compression does not help the snap of percussion instruments, but it can increase their punch.	If you don't like the percussiveness (e.g. snare drum), slow down (lengthen) the attack. To increase the ratio of rhythmic snap to smoothness, shorten (speed up) the attack. Upward expansion is very good at helping the snap of percussion instruments, however, sometimes at the expense of the vocal balance because percussion becomes more prominent. Too much expansion and punch is lost.
can work very well with upward expansion, which can enhance some transients which compression may have overly softened.	can work very well with upward compression, which fills in any perceived low level "holes" or lost sustain.
Very easy to degrade the liveliness or "bounce" of the music if time constants are not optimized or if overused.	Very easy to enhance the liveliness or "bounce" of the music, but watch out for too much "bounce" or exaggerated dynamics.
can decrease the overall dynamic range of the song (macrodynamics), in addition to affecting the microdynamic impact of the music.	can increase the overall dynamic range of the song (macrodynamics), making a climax seem even more climactic, which can be very effective.
tends to de-emphasize musical accents and emphasize the sub accents and sustains in reverse proportion to their original movement.	tends to emphasize the hottest musical accents and to a lesser degree, the sub accents in increased proportion to their original movement.
	can be followed by a limiter to prevent loud (expanded) passages from overload. As long as the limiter is used to cheat down very short, momentary transients, it will not significantly diminish the effect of the upward expansion. The limiter's gain reduction meter should be moving very little and on brief occasions, while the expander's gain increase meter should be bouncing with the syllables of the music that's being enhanced. Be careful that the limiter does not audibly diminish the benefit of the expander or there is no point.

CHAPTER 8

Audio Restoration

I. Introduction

In this chapter, we introduce the subject of audio restoration, largely concentrating on noises and distortions that we may encounter in modern recordings. I could write a whole book on this topic, which would include important practices like restoring from LPs when the tape masters are not available, or transferring and restoring analog tapes, or 78 RPM records, areas not covered in this book.

Noise and Distortion

Specialists in any developed subject create a vocabulary to mark distinctions that are not generally noted, or even noticed, by the non-specialist. Although a layperson would lump distortion and noise together, an audio engineer characterizes distortion as a particular form of noise: one that is correlated with the signal. Distortion can be low level and sound much like what is normally called noise, or it can be high level and quite obtrusive, lying on the peaks of the signal.

Continuous or Impulsive

Noise can be either *continuous* (with little or no dynamic movement), or *impulsive* (intermittent or periodic). Some examples of *impulsive* noise include: crackle, click(s), tic(s) (very short duration clicks), and pops (primarily low frequency). *Continuous* noise is further divided into two categories: *broadband* and *tonal*. What distinguishes broadband from tonal noise is that although the former can have a frequency response character or *color*, it has no obvious identifiable single frequency component(s). The color names we use to describe broadband noise include *white* (wideband with a rising high frequency response), *pink* (wideband with a flat frequency response), "It takes a lot of noise to distract the average listener, but not much noise reduction to harm the sound."

rumble (narrowband with a distinctive bassy character), or *hiss* (narrowband with strong components in the 2 to 10 kHz range). By contrast, tonal noise contains distinct components at single (or multiple) frequencies, e.g., feedback, buzz, or hum. **Hum** consists of the lower frequency components of the power line. In Europe and Asia the powerline fundamental is 50 Hz, with lower harmonics of 100 and 150 Hz; in the U.S. it's 60 Hz, with lower harmonics of 120 and 180 Hz. **Buzz** consists of the higher harmonics of the power line frequency, running (in a 60 Hz series) from 240, to 360, and right up to 2400 Hz and higher in severe cases.

Why Reduce Noise?

The first type of noise that often comes to mind when we think of noise is *tape hiss*, a potential problem when restoring old analog tapes. Preamp hiss from musical instruments and microphone amps is less prevalent today, but still found on modern recordings. When the CD first appeared, some mastering engineers were overzealous about removing noise, because the medium was billed as silent — but this practice resulted in botched masters with ugly artifacts: silence where there should be room tone, loss of ambience and definition, and noise modulation. Fortunately, we soon became less concerned about hiss from analog tape sources. Listeners have become used to the idea that a classic analog source master might be a little noisy, and engineers recognize that noise reduction has its tradeoffs. In fact, it takes a lot of noise to distract the average listener, but not much noise reduction to be debilitating to the sound. Keep this in mind when deciding when to attack noise and when to leave it alone. When the noise is *intermittent* (as opposed to *continuous*), it is far more problematic because it attracts attention when it comes in and out.

A lot of project studio mixing rooms are not as quiet as they should be: air conditioner rumble, airflow noise, and fans in computers cover up noises in the mix. Regardless, the mix engineer should be concentrating on other things than whether or not the singer produced a mouth tic. Consequently, when the mix arrives at the quiet mastering suite, we notice problems that escaped the mix engineer — mouth noises, click track leakage, electrical noises, guitar amp buzz, preamp hiss, or noises made by the musicians during the decay of the song. We use our experience to decide if these are tolerable problems, or if they need to be fixed: the criterion is whether the noise would distract the listener from the pleasure of the recording. Hiss traceable to a single instrument track is more transparently fixable at mix time; I would ask the mix engineer to send me the offending track for cleanup, then return it cleaned for the mix. Or I might suggest a remix, bringing his attention to vocal noises that he can mute. But clients don't always have the luxury or time to remix, and so mastering houses need to have the most advanced noise reduction tools that will affect the surrounding material as little as possible.

Noise Reduction Processors: A Partial List

This list includes mature products that I have used, or that have been recommended by trusted colleagues.

- Stand-alone (outboard): Cedar Cambridge (every style of noise reduction plus other features such as linear-phase EQ); GML 9550 (broadband denoiser); TC Backdrop (broadband and tonal noise reduction for the System 6000); and Weiss DNA-1.

- DAW-integrated: Cedar Retouch and Algorithmix Renovator, customized to each DAW, as is NoNoise (available in soundBlade or Pro Tools); Sequoia and Wavelab have integrated noise reduction tools.
- A la carte plug-ins (AAX, RTAS, VST): Products from Algorithmix, Cube-Tec, Izotope, and TC (Powercore, discontinued but still available).

The noise reduction methods described in this chapter are **single-ended**, as opposed to **complementary**. Single-ended systems attempt to separate noise from signal without having a specially-recorded track, whereas a complementary, or two-step, noise reduction system, e.g. the Dolby® system, applies one process during recording and an equal and opposite process during playback.

The Remedies

Each kind of problematic noise requires its own dedicated technical cure. Sometimes the best cure is just to ignore the noise! We engineers tend to forget that the ear has a built-in noise reduction mechanism that lets us separate signal from noise and hear information buried within the noise. Thus the key to effective noise reduction is not to attempt to remove all the noise, but to accept a small improvement as a victory. Remember that louder signals mask noise, and that the general public does not zero in on the noise as a problem. They're paying attention to the music, as should we!

LP transfers and other sources may contain hiss, hum, rumble, crackle, clicks, pops, tics, and peak-level distortion. Each type of noise or distortion requires a dedicated correction algorithm: for example, when noise is continuous we choose either a *broadband* or *tonal* processor. Broadband processors can be further tailored to work in selective frequency ranges such as rumble or hiss or by manipulating any number of band thresholds or equalizing the noise reduction curve. Some noises that seem to be continuous – for example, air conditioner rumble – actually contain periodic (repetitive) components that may need to be treated as impulsive. On the other hand, impulsive noise cannot be fingerprinted, so our algorithm must concentrate on known characteristics of a click, for example, and set a semi-automatic threshold. Impulsive noise reduction systems specialize separately in clicks, pops, crackle, scratches, etc. The dividing line between *clicks* and *crackle* is fluid; crackle consists of many closely spaced clicks. Sometimes crackle can be removed with a declicker, and vice versa. Cedar's dedicated *Declickle*™ algorithm distinguishes clicks and crackle from signal. Sometimes a denoiser that requires a *fingerprint* (see below), normally designed for continuous noise, can perform excellent decrackle or debuzz.

Fingerprint or No Fingerprint?

A **fingerprint**, also known as a "noise profile," is a sample of noise without signal (even one second will do). Usually I loop the sample for several seconds and feed it into the processor, which learns its characteristics and develops the antidote. If a sample is "contaminated," that is, if it contains some signal, then the noise reduction will be less effective or have artifacts. Some noise reduction units either require or can optionally use a fingerprint. Systems that do not employ a fingerprint have become increasingly effective. They are very smart **noise gates**, usually threshold-based, and they operate either in selective bands or across the entire spectrum, using an equalized response to avoid affecting signal while reducing noise. The skill of the operator is far more important than how the noise-reduction system works. I prefer to use a fingerprint method when a good noise sample is available, but keep the other system around when one is not. You must learn what the best tool is for each job.

II. Noise Reduction – Simple to Complex

When it comes to denoising, each processor has its specialties, learning curve and artifacts. Ease-of-use is usually, but not always, inversely proportional to effectiveness: consumer software, for example, with its simple setups, provides the least satisfactory results. If you require effective denoising with the least amount of artifacts, spend some serious time learning how best to use your tools. There is no single best tool for all jobs, and the best tools do not always come from one manufacturer. Everyone is initially seduced by noise-reduction power, until the artifacts begin to call attention to themselves. Each practitioner has his priorities: some value power over sonic transparency, and vice versa. For example, preserving depth and space is more important to me than having a perceptibly silent background.

Simple Equalization

The simplest cure for hiss is an equalizer, as long as the passage does not contain instruments playing in the hiss range, or whose harmonics cover the hiss range. For example, an electric piano solo introducing a song may be hissy, but that noise will be masked when the rest of the instruments enter. This is a candidate for a filter active only during the piano introduction; say 1 to 4 dB dip around 3-5 kHz (this is the range where the ear is most sensitive to hiss). We have to trade off the loss of piano harmonics versus the improvement in noise. Hiss may draw attention to itself during a decay. In Sequoia, using object-based processing, it is simple to apply an equalizer unobtrusively, just on the decay.

P-pops are a signal-related noise, so they are a form of distortion. And since they are primarily low frequency, they can be treated with a selective high-pass filter, up to about 100 Hz. As long as the filter is applied briefly, the result can be artifact-free. Using SADiE, I would capture a short section with the filter, then, using the crossfade editor, narrow the extent of the filter to the p-pop. This edits out just the offending portion. With practice, the technique can be extremely fast. In Sequoia, I can make a cut to isolate the p-pop in an object and apply a non-destructive high-pass filter just on the object. But the most direct way to fix a p-pop is with an integrated spectral editor (see below).

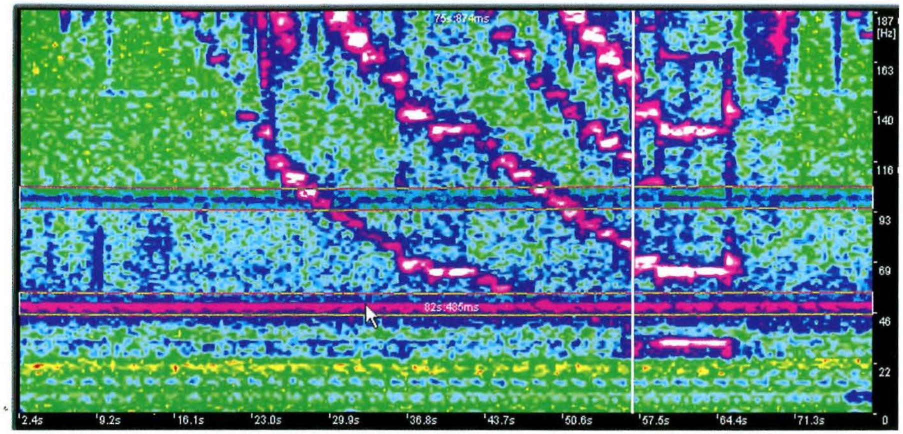
Narrow-Band (Downward) Expansion

Compression techniques used in mixing and mastering can bring up noise in original material from tape, preamps, and guitar and synth amplifiers, all of which could be problematic. Since compression aggravated the noise, expanders are its cure. As little as 1 to 4 dB of noise reduction in a narrow band centered around 3-5 kHz can be very effective. These units typically have 3 to 4 bands, but we will use only one. Start by finding a threshold, with initially a high expansion ratio, fast attack and release time. Zero in on a threshold just above the noise level. You'll hear ugly *chatter* and bouncing of the noise floor because the time constants are so fast. Now, reduce the ratio to very small, below 1:2, perhaps even 1:1.1, and slow the release until there is little or no perceived modulation of the noise floor. Too much expansion, and you will hear artifacts such as pumping or ambience reduction — or the expander will interfere with the music itself. The attack will usually have to be much faster than the release so that fast impulses will not be affected. Depending on the music, its dynamic characteristics, and its original SNR, this subtle approach can yield artifact-free noise reduction. The other expander bands should be bypassed or have the ratios set to 1:1. An expander's look-ahead delay (see Chapter 6) allows it to open before the signal hits it, thereby conserving transients. If the expander approach does not work, then we have to apply more sophisticated, dedicated noise reduction processors. A *noise-gate* is an expander with a very high ratio. Personally I find the noise-gate's cure worse than the disease.

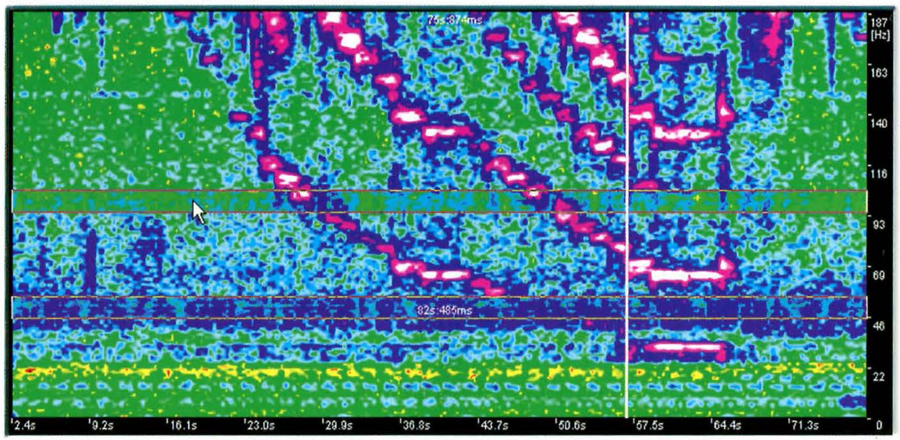
Complex Filtering for Tonal Noise

Now we begin to look at processors dedicated to the task of removing noise. Tonal noise can be reduced using narrow-band selective filtering. The practical limit of Q is from 40 to about 100, before filters produce ringing artifacts. *Sonic Solutions No-Noise* (available for Pro Tools and soundBlade) has a complex filtering option that lets you insert many high-resolution narrow-band filters suitable for removing hum and buzz. Before inserting the filters, it's useful to do an FFT analysis of the noise floor to see which harmonics are present so you can apply only the filters that are needed. Sequoia's phase-linear FFT filter can do a similar job. SADiE has enough DSP power to insert many narrow-band filters in real time; my dehumming preset has about 25 filters set for a Q of 40 or higher. I selectively bypass each filter to hear if it's needed, and set its reduction just enough to reduce the tonal component below the annoyance level. TC's *Backdrop*, normally used for broadband noise, has a preset which, with a fingerprint, can be very effective on hum and buzz.

Izotope RX and Renovator both have a harmonics removal tool. Renovator's selection mode can pick out odd, even or both sets of harmonics in a selectable frequency range. The interface is so ergonomic that what used to take hours using a text file takes only a couple of minutes with Izotope or Renovator. The above images show a simple and quick fix for 50 and 100 Hz hum.



Complex filtering in Renovator. Hum at 50 and 100 Hz has been selected.



After processing, the hum is gone with little or no effect on the musical notes. The selection lines remain to show the areas where the processing had occurred.

Buzz removal is very difficult to do successfully, and requires a different (stronger) algorithm than that for treating hum. Cube-Tec has the most effective debuzzer I have tested. It has both a learning (fingerprint) mode and a manual mode. It concentrates on harmonics of a defined fundamental (usually 50 or 60 Hz) and lets you

Cube-Tec Debuzz Plug-in

select how many harmonics it will affect. It can also follow the fundamental if it varies due to speed variations or even flutter (pictured above).

Broadband Processors

Dedicated broadband noise reducers are sophisticated multiband downward expanders with many bands so that one band's action does not affect the other, to reduce artifacts. Those that can use fingerprints calculate the expansion threshold of each band, which can then be fine-tuned by the operator. *Algorithmix NoiseFree*, *Cedar Denoise*, *Izotope RX's denoiser*, and *Sonic Solutions NoNoise* all work best with fingerprints. The task of finding a fingerprint can be made easier when the client sends in samples of the noise with no music playing; thus, when sending material in for noise reduction, the mix engineer should not tightly cut the beginnings of material; the noise just before the downbeat is an excellent candidate for a fingerprint. By manipulating thresholds or the bands of a broadband processor, a skilled operator can tailor the frequency response of the noise reduction curve for the best compromise between artifacts and perceived noise reduction.

Systems that do not use fingerprints separate the signal from the noise by either using an automatic algorithm, or by giving the operator manual control over the threshold of each band or range; Weiss's standalone DNA-1 is one such unit with algorithms for broadband and impulsive noises. Izotope's RX noise-reduction plug-in does not require a fingerprint. There are two standalone units designed for real-time manipulation by the operator: the Cedar DNS1000 and the GML 9550.

For effective broadband noise reduction, try to find the earliest generation musical source. Second generation analog tapes contain two layers of different noise, and can easily confuse any processor.

Declickers

When using declickers, test for artifacts with the *difference* button, which plays the difference between source and output, so ideally all you should hear is the unwanted click with no evidence of the wanted signal. Aggressive automatic declicking can distort the peaks of high frequency instruments such as trumpets. If there are artifacts, lower the sensitivity and try again, or replace compromised portions of the auto-declicked file with the source, then use surgical manual declicking.

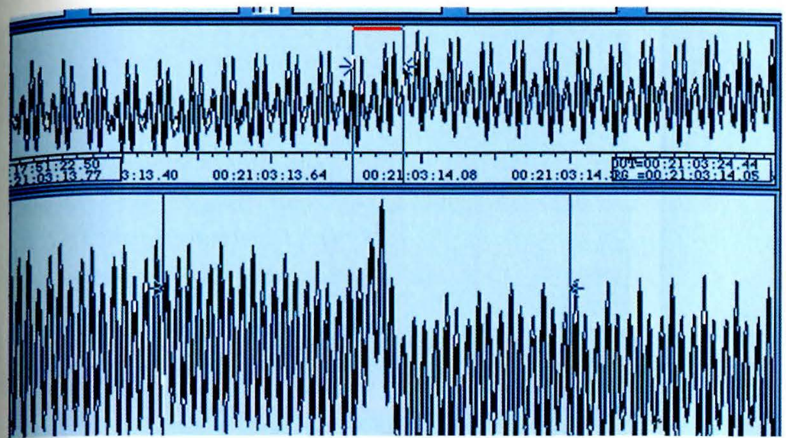


Figure A *LP Thunk in Sonic solutions. Left channel, area indicated by red bar has been denoised. Right channel has not yet been processed. Different panel heights reflect different visual magnifications, not different amplitudes.*

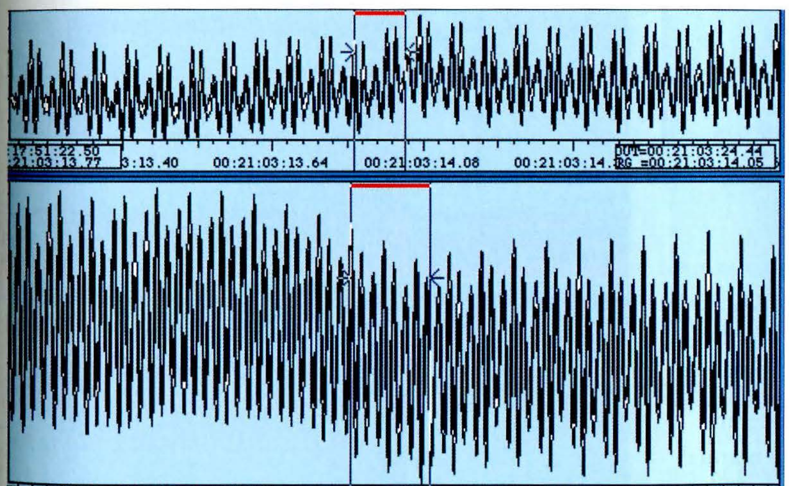


Figure B *After manual declicking, the right channel thunk has been removed.*

The pictures at left illustrate the power of Manual Declicking. Figure A shows a "thunk" from an LP record. The left channel (top panel) has already been dethunked, as can be seen by the horizontal red marker above the waveform. When reproduced, the slight DC level shift that remains does not translate to an audible noise. The right channel contains a severe thunk manifested by an instantaneous upward, then downward, DC level shift (that will cause woofers to rattle). With Sonic Solutions manual declicking, the correction process is as simple as marking the noise with the gates and selecting *D Type* from the menu. D Type is a powerful interpolator which can stitch together "impossible" waveforms and even remove brief dropouts or holes with no audible artifacts. In Figure B, the low frequency thunk and most of the DC discontinuity have been repaired; the ramped DC level shift that remains (probably record warp) does not produce an audible noise.

MYTH:

"I know that you can't hear anything but noise on this tape, but if you get rid of it all, you'll be able to hear my husband having sex with his lover."

CONTRIBUTED BY GORDON REID (CEDAR)

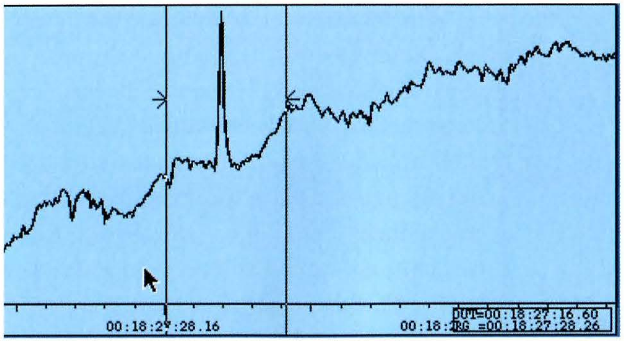


Figure C

A click is surrounded by the gates.

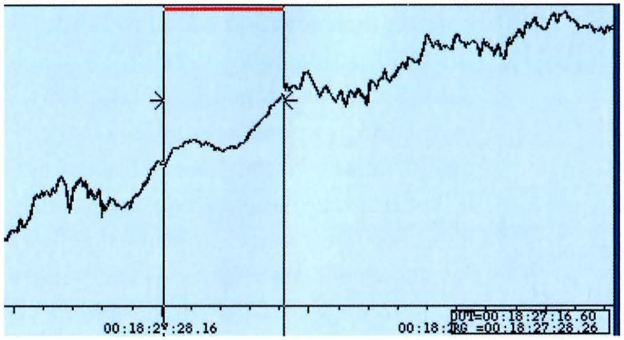
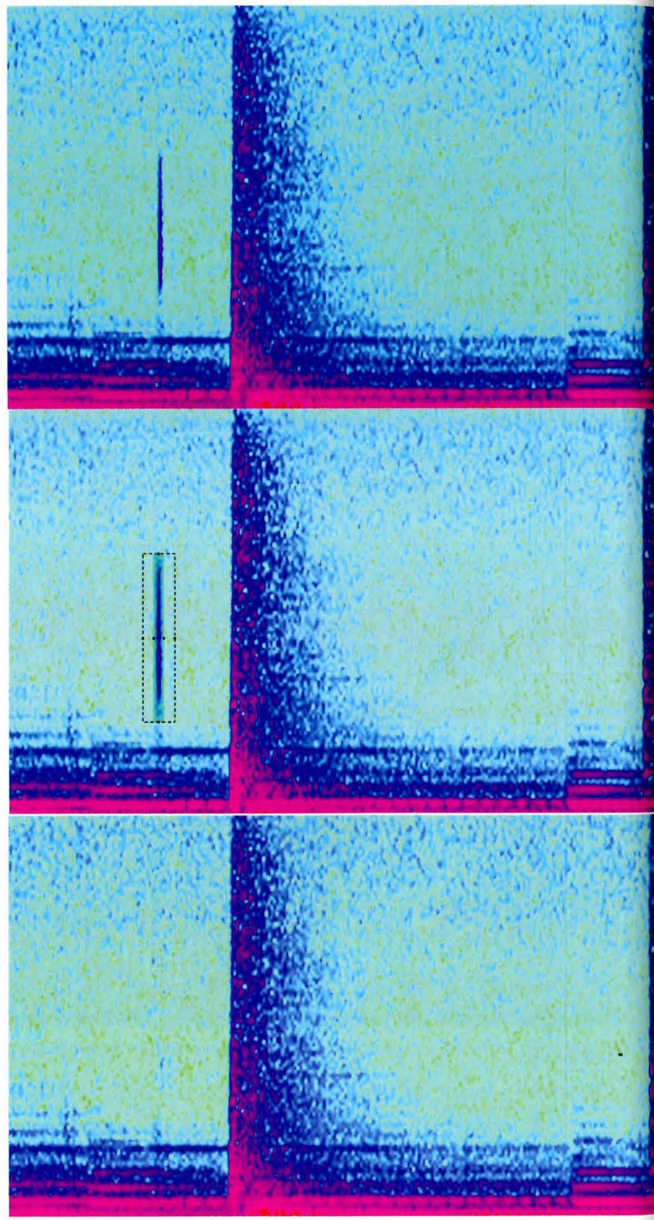


Figure D

After choosing from the No-noise menu, the click is removed (marked by red bar).



Cleaning up mouth tics in Retouch is as easy as 1,2,3

In Figure C (above) a severe click is marked manually by the gates, and in Figure D it has been removed. Note that Sonic Solutions' automatic vertical gain conveniently amplifies the display to the highest amplitude in the view.

Spectral Editors/Processors/Interpolators

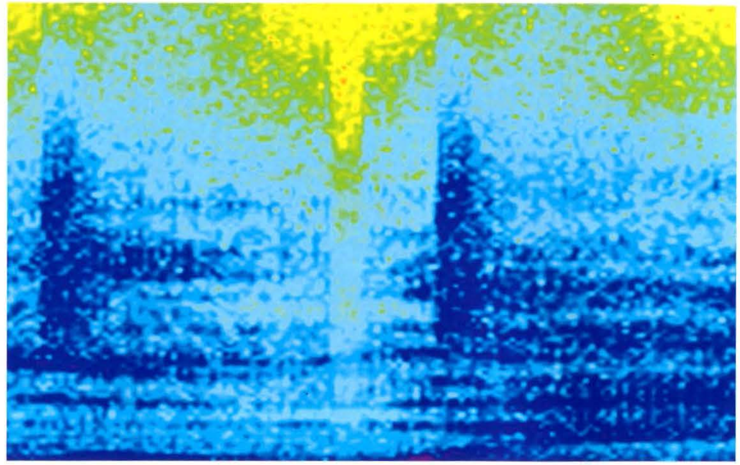
Spectral editors have become essential mastering tools, superseding and performing better than manual declickers, some denoisers, and pop-removers. They can very transparently remove noises that were previously unfixable, such as a baby crying, chair squeaks, even people talking in the middle of a musical take! Compared to older audio repair technologies, they can be so surgical, there are usually no audible artifacts. The simple procedure is to make a selection around the noise. The processor then replaces the noise with information from the area surrounding this region, using an interpolation algorithm to perform a smooth transition between the original and the corrected sound. These tools must be *integrated* with the DAW in order to be most effective: The engineer creates a range or an object which contains the noise, then calls a keystroke to launch the processor, cleans up the noise and seamlessly replaces the object with the cleaned version. Currently these processors are in flux. I like the power of Renovator, but it does not function in 64-bit Sequoia. I like Retouch's ease-of-use, but it is only available in the Pyramix and SADiE DAWs. I think Sequoia's tool is very good but not quite as effective or easy-to-use as the others. Izotope's Spectral tools are very powerful, but for integration we have to wait for a future version.

The spectral display looks at music three-dimensionally: frequency from top to bottom (read like a musical score); time from left to right (also like a musical score); and intensity expressed in colors from dark through bright. Its power can be seen in the three figures (page 118) taken from Retouch; cleaning up mouth tics is as easy as 1,2,3.

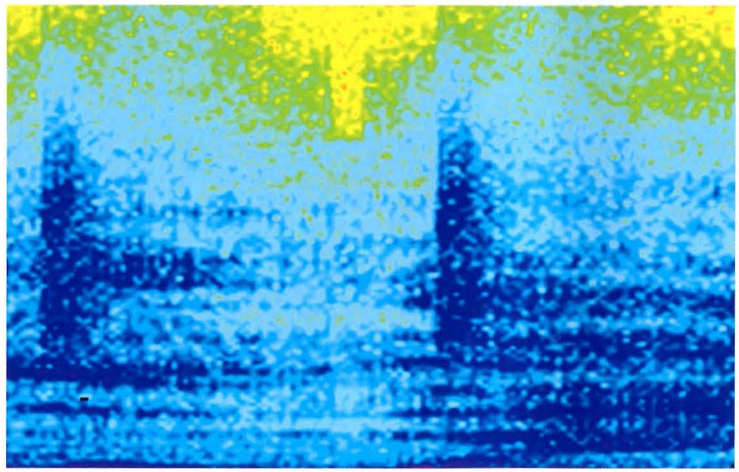
1) Locate the tic, easily seen within the surrounding music. This tic occupies the upper midrange to high frequency portion of the spectrum.

2) Draw a box around the tic. Retouch then adds a dashed rectangle, indicating the area that will replace the tic.

3) Press the **Retouch** button. The tic instantly disappears, damaged material replaced by surrounding information from the left and right side. This process is called *interpolation*.

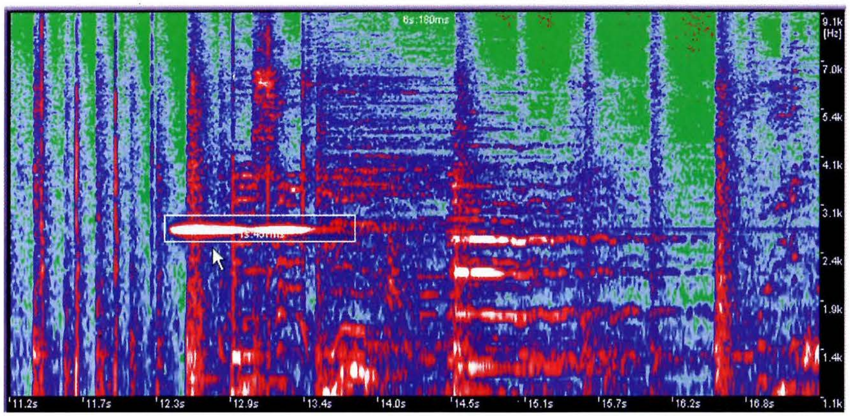


A dropout in an analog tape (located between two major beats of the music).

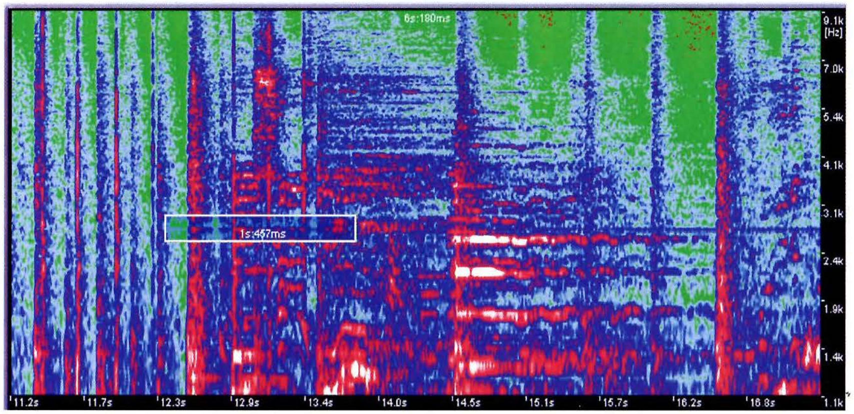


After Retouch, the hole is gone and the sound completely restored. Whatever visual remnants that remain of the hole are audibly masked.

Spectral Editors can deal with the absence of sound just as easily as they remove an annoying one — for example, fill in holes caused by analog tape dropouts (illustrated above, example from Retouch). Sometimes a special mode is required to tell the interpolator to fill in holes; in the case of Renovator, we select a gain of -60. Additionally, I've used spectral interpolators to help construct seamless room tone, to soften or reduce distortion, and as a manual de-esser. I've used the patching tool



Acoustical feedback has been identified by drawing a box around the offending sound.



The feedback has been removed. Intepolation occurred vertically, from other frequency elements played at the same time.

to edit music, replacing sections over damaged pieces of music. For p-pops, these interpolators are much faster and more selective than the old method of a high-pass filter followed by DAW editing.

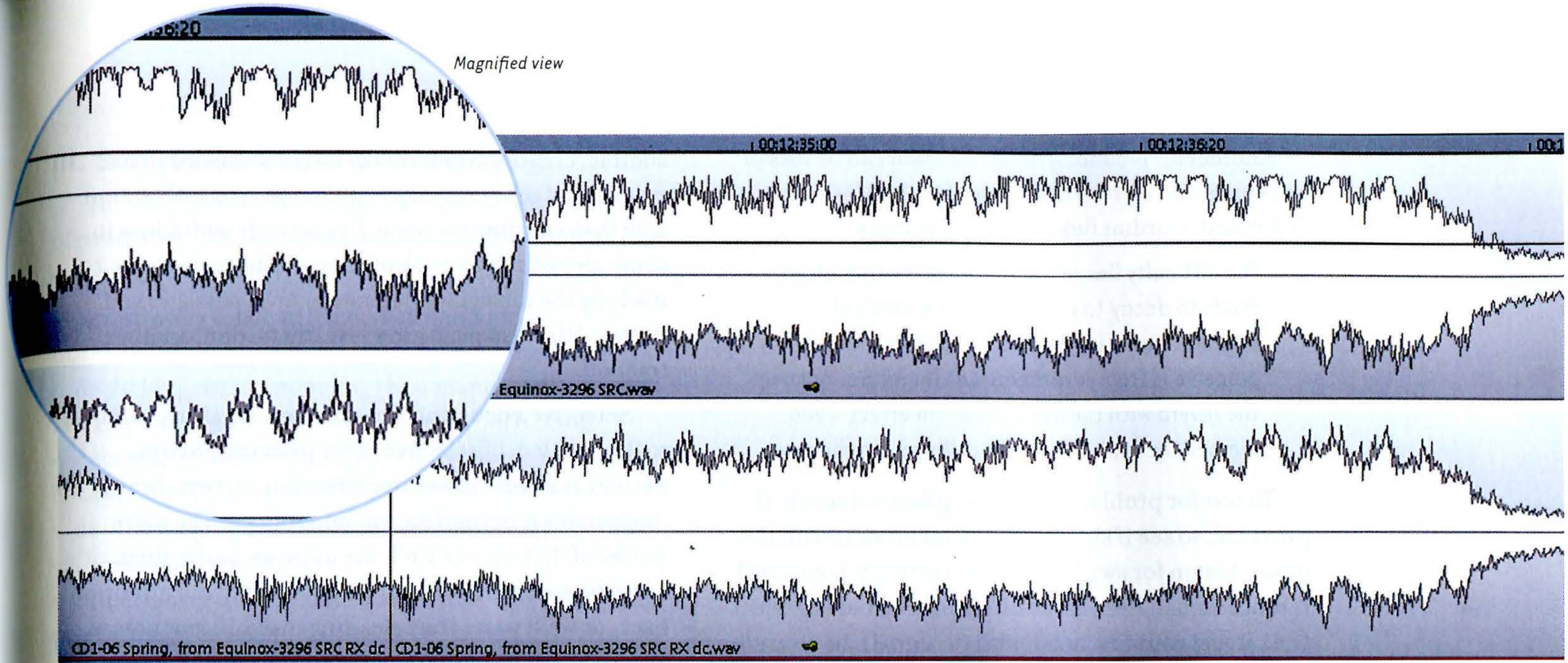
Next is an example taken from Renovator (pictured above). Acoustical feedback seen in the first image is instantly gone in the second. In this case I interpolated vertically, from frequency information above and below the feedback.

Distortion Removal

The simplest solution for short periods of distortion is to disguise it: spectral editors used in gain mode instead of interpolate mode can selectively soften high frequency regions, reducing the harshness of distorted passages with few noticeable artifacts. A **decrackler** or **descratcher** can make an excellent distortion-softener or remover, when selectively applied. Sonic Solution's *E Type* **manual declicker** is a good fixer for very short overload distortion. These algorithms replace the distortion with an interpolation to approximate the original sound.

Clipping that is level-dependent and has a clearly-defined threshold is the easiest to repair. Dedicated **declippers**, such as Cedar's *Declip* and Izotope's **Declipper** remove (or at least reduce) clipping distortion by interpolating the missing pieces. When using a declipper, be sure to drop the input level enough to leave room for the restored material, which has a higher peak level. If possible, store the output of the declipper as a 32-bit floating point file in case you discover a clip later (Chapter 16 will explain floating point file formats). Be careful when setting the threshold of declipping: more is not better! In fact, it is easy to cause distortion instead of removing it by setting the threshold too strongly.

The next figure (page 121) is a remarkable illustration of Izotope's Declipper. Notice that the clipping in this case was only in the positive direction. The music is an orchestra accompanying a brass choir, which sounds horrible when clipped, probably caused by a preamp or ADC overload. Notice the visual restoration of transients and microdynamics; the sound improvement is as impressive as the picture.



Brass choir: original, clipped (top image). Declipped with Izotope RX2 (bottom). View of left channel only.

Other Specialized Processors

When a tonal noise is varying in frequency, as in the case of analog tapes with varying speed, a special kind of **tracking filter** is required, usually found in forensic suites. *Dethump* from Cedar is dedicated to long low-frequency thumps and scratches. A *Deplop* algorithm (available from Cedar and Cube-Tec) handles the low-mid frequency ringing artifact that may remain after a click is removed. *Phase, time, and azimuth* correctors are available from Cedar, Cube-Tec, and Izotope, though it's still very important to get the **azimuth** right during the tape transfer. To be an effective azimuth corrector, the tool must be able to perform *subsample time delays*, done by first upsampling, then downsampling. Be wary of automatic azimuth adjusters, which can mistake normal microphone delays for phase shift and ruin the stereo depth of a recording. With some tracking filters, you can engage the automatic correction, then freeze the adjustment so it will not wander, and manually tweak after the automatic correction has taken place. As always, listen. Izotope has a **dereverberator**. I haven't tried it yet; be sure to listen for artifacts since devices like these can produce the **space monkeys** (see below).

Artifacts and Perspective

No single-ended noise reduction system is perfect; all noise reduction systems remove some signal along with the noise, and may add noises of their own. Ironically, a quieter original recording can be more effectively processed, because the more separated the original signal is from the noise, the more easily the noise reduction system can operate without hurting the signal. So a really noisy recording probably cannot be fixed without creating artifacts. Artifacts of overaggressive denoising include: comb-filtering, swishing or phasing noises (known semi-affectionately as *space monkeys)*, and low level thumps and pops (that can be worse than the disease).

Another *by-product of noise reduction* can be loss of ambience and stereo separation. On the Mastering Webboard, Gordon Reid of Cedar explains:

> The difficulty lies in the fact that reverberation tends to decay to noise. However, much of the directional information and ambience we perceive is from reverberation. Therefore, remove the reverb with the noise, and – in effect – you remove the walls, floor and ceiling from the room.

To test for problems, use the *difference* button, if provided, to see if signal is being taken away with the noise. Listen for swishing noises (artifacts) in the difference signal. Even if the difference signal is perfect (i.e., it just contains noise with no signal), be aware that, psychoacoustically, the presence of noise increases apparent high frequency response. So remove only the annoying portion of the noise, and be prepared to restore any high end that seems to be lost. Or, try my own *K-Stereo* processor to restore lost ambience and depth due to noise reduction.

The Layers of the Onion: What distinguishes good noise reduction work from bad is finding the optimum amount, because as noise is removed, more noise is revealed (noise itself masks other noise below it)! Beneath each layer of the onion is another layer. If you remove hiss, you may then hear crackle that was not previously audible, creating a potentially larger problem, since crackle and other impulsive noises are more objectionable than continuous noise. I've actually *added* hiss to some recordings to mask objectionable movements made by the musicians during low level passages and decays. Hiss also masks low level distortion, another tradeoff.

Selective Focus and Perspective: We all succumb to the problem of selective focus from time to time: as soon as a noise draws our attention, our ear-brain exaggerates it beyond its real importance. The more noises of that type we find, the more we notice them. It's a psychoacoustic problem. We have to pull our attention back, or we'll waste time cleaning insignificant noises. Be aware that headphones exaggerate noise. Another consideration is the client's perspective. I once mastered an album in which the opening of a tune had an obvious electrical tic on top of the bass player's note. I removed the tic, restoring the note to its beauty, I thought. But then the producer asked me to bring the tic back – demonstrating that many noises are considered to be part of the music. Become familiar with each musical form – sometimes "dirty" is "clean".

In all cases, careful judgment is required to ensure that the music has been better served.

III. Suggested Order of Processing

To minimize artifacts and deal with the interaction of processes, it is best to treat noise in a particular order (each step is optional):

- Remove any tonal artifacts that stand out (e.g. hum, buzz), using a simple or complex filter, followed by
- declicking, first automatic, then manual to deal with any remnants not caught by automatic declicking
- decrackling (which can also remove some remnant clicks)
- de-clipping or distortion reduction
- broadband denoising
- Finally, overall program equalization, filtering, other processing if needed

Each successive process should be saved to a new file including the names of the previous processes used. For example, "The Look of Love fl.wav" is the filtered file. "The Look of Love fl+dc.wav" was first filtered, then declicked. If using a floating point DAW, save the intermediate products in floating-point format.

Audio restoration has become far less labor-intensive with the inventions I've described in this chapter. It's still work, but very rewarding – like hiring a meticulous gardener to remove each weed in your garden by hand, instead of using harmful chemicals.

"No single-ended noise reduction system is perfect; all noise reduction systems take away some degree of signal with the noise."


SONY
Pendulum Audio
DS1-MK3
DS1-MK3
COMPRESSOR
AnaMod
ATS-1

CHAPTER 9

Additional Mastering Techniques

This chapter takes us from required basics to advanced mastering techniques including how we can "make it louder" with the least compromise (if the producer requires a "hot" master).

I. Basic "Objective" Techniques

Mono Check for Loudspeaker Integrity

Before we begin mastering, during the first listen of the day it is a good idea to check our loudspeakers by putting the monitor into mono and playing a wide range musical selection. If the center image is tight and unwavering, this confirms that left and right channel loudspeaker drivers are matched in frequency response and level, that a tweeter or crossover component has not gone bad on one channel. If we suspect a problem, then take proper measurements as described in Chapter 21.

Stereo Balance of the Program Material

Music feels much better when the stereo balance is "locked in," which can mean as small as an 0.2 dB level adjustment in one channel. It is generally unhelpful to use meters to judge channel balance because at any moment in time, one channel will likely measure higher than the other. I've seen songs where one channel's meter is consistently a dB or so higher than the other, but the balance sounds exactly correct. Centered vocals are a good indication, but there are always exceptions. Proper balance should be determined by ear; I never use a so-called stereo position indicator because our brains and ears do a better job. Assuming that the vocal is supposed to be centered, you can test with a mono switch which puts everything in the middle. When you do this, the vocal position should not shift.

Thor Legvold has offered this tip: Use *Left/Right swap* to check stereo balance. Centered vocals stay right in the middle, while the rest of the band trades places. However, vocals are not always centered, and even if the lead vocalist is centered, the result may not be the optimum left/right balance for the particular piece of music. Listen carefully, and try to determine the producer's intent and the musicality of the balance. If the vocal is off-center, listen to the side instruments to see if the producer did this intentionally to help the overall stereo picture and mix balance.

Fixing Interchannel (Relative) Polarity Observing Phase Shifts Between Channels

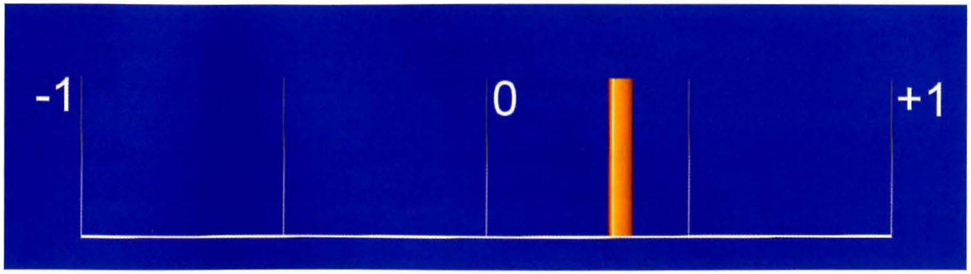
Phase is a concept that operates within the time dimension. The so-called *phase switches* on consoles are misleading because they *do not shift phase*, but instead invert signal *polarity*. If two sources, especially left and right channels, are 180° out of phase at all frequencies (or a large band of frequencies), we say they are *out of polarity with each other*, and so the polarity of one of the channels must be corrected to compensate. A polarity error can impart a hollow quality to the stereo image, reduce the bass response, and even move the image behind the listener. In mono, everything will cancel if there is a complete polarity reversal between left and right channel. If some instruments in the left channel are out of polarity with those in the right, then those instruments will cancel in mono. In fact, the best way to test for interchannel phase or polarity issues is to listen in mono. Listen for loss of bass, lost instruments, or other cancellations. Invert the polarity of one channel and see if the sound is better or worse in mono. Stereo recordings made with spaced omni microphones may exhibit ambiguous phase on a correlation meter (to be explained), but usually listening in mono will reveal the one correct polarity setting. At the mastering stage we can only correct a relative polarity issue if the entire mix has a problem. If, for example, the percussion drops out in mono but the vocal remains, then a remix is required. *Phase shift*, as opposed to polarity, is time shift: e.g. if the snare drum is not equidistant to both overhead mikes, this can adversely affect the frequency response of the snare, especially when auditioned in mono. When mixing, phase and polarity issues are important between microphones that were captured at the same time, e.g. a multimiked drum set or an acoustic piano. However, the polarity of a single-track overdubbed solo instrument doesn't matter at all, unless you have simultaneously overdubbed two or more tracks, e.g. a direct box and the microphone. In this case, timing (phase) and polarity between these two elements are very important.

How to Use a Correlation Meter: A correlation meter, also known as a phase meter (pictured on page 127), can tell you that something is amiss, and even indicate that some information in the left channel is out of time (out of phase) with that in the right. The meter is marked from -1 at the left through 0 in the middle and +1 at the right (sometimes 180° at left, 90° in the middle, 0° at right). 0 means "random correlation," meaning that the left and right channels are only randomly related to each other. A 0 reading can occur when the sound consists of stereo ambience, decay, or if there are two entirely different programs in the channels. +1 means the sound is completely mono, and that there is 100% correlation at all frequencies between the left and right channels. A constant reading of -1 means the channels are correlated, but one channel is out of polarity with the other. In that case, it is also correct to say that all elements in one channel are 180 degrees out of phase with all elements in the other, at all frequencies. If the channels were partially out of phase, the meter would be to the right of the -1 position. Meter movement towards the middle is desirable: it tells you there is a lot of random phase information in the stereo field and that the stereo image will likely sound rich and spacious. If the meter is slightly off of its extremes, you'll know there is some phase shift between the channels. But don't try to correct timing information unless *all* the information in the left channel is out of time with that in the right.

If the high frequency response of all the instruments gets worse when listening in mono, it is an indication that there may be some phase shift between channels. If so, we may improve the situation by adjusting inter-channel timing with an azimuth-correction plug-in (See Chapter 8) while listening in mono. Cedar's digital azimuth corrector makes timing adjustments in sub-sample increments. It is accurate to 1% of a sample. A similar procedure is used by engineers to align spot microphones with the main mikes (see Chapter 10).

Is It an Innie or an Outie?

In the real world, some musical instruments create asymmetrical waveforms. The direction of pressure of the wave should be preserved from the recording microphone to the consumer's loudspeaker and the listener's eardrum. The standard *AES26-2001* states that microphones must produce a positive-going voltage on pin 2 when excited with an **acoustic compression** — an increase of the instantaneous sound pressure that causes displacement of the microphone diaphragm away from the sound source. This should be represented by an upward-going waveform in the DAW, and on replay translate to a displacement of the loudspeaker diaphragm toward the listener. Then we can state that the reproduction system has correct **absolute polarity**. Absolute polarity applies to all program channels.



Correlation Meter reading almost 3/4 of the way towards the right, which indicates only moderate stereo separation and likely a fairly narrow stereo image.

To confirm that the reproduction system has correct absolute polarity, we can use a device called a **popper**, which sends an asymmetrical signal into the system and measures the output of the loudspeakers with a test microphone. You can connect a 1.5 volt battery momentarily to the speaker terminals, then observe the direction the woofer cone moves, which I call a "pauper's popper." With active loudspeakers, connect a battery momentarily to pins 2 and 3 of the XLR, positive on pin 2. To avoid loudspeaker damage, start with the monitor control turned down and turn it up only far enough to be able to see the woofer's movement. If the woofer is hidden behind a grill, try to shine a light on it through the grill. Then, if you have a separate subwoofer and satellites, feed wideband pink noise to the loudspeakers and invert the polarity of either the low or high frequency portion. The polarity combination that produces the loudest bass is the correct one, meaning that the two parts of the system are in phase with each other. After that, send a momentary positive pulse from the source DAW and confirm that the woofer moves outward. This will verify the polarity of all the audio connections between the DAW and the loudspeaker.

Once the monitor system has been verified, you can check absolute polarity of the program material by reversing the polarity of both channels in the DAW. Do you hear a difference? On some systems the difference is extremely subtle, but it is possible to detect absolute polarity differences on certain highly coherent loudspeaker systems.[1] I produced an absolute polarity test for Chesky Records, using a solo trumpet recorded in a natural space with a Blumlein microphone pair. When the polarity is incorrect, the trumpet appears (to most listeners) about a meter further back. This is evidence that incorrect absolute polarity can affect how we mix and master.

The DAW waveform is deceiving: most times we cannot determine the absolute polarity of an instrument by looking at its waveform, since not every instrument begins playing with the first sample going "up." Stick with listening. Experiment with both polarities and judge which sounds better.

DC Offset Removal

Sometimes poorly-calibrated ADCs or poorly-implemented DSP processes can add a DC offset, which means that the centerline of the waveform at rest is not exactly 0 volts. When the offset is excessive, overall headroom is reduced; raising gain in this case would cause the audio to clip prematurely in either the positive or negative direction. The best way to determine if a DC offset is a problem is to repeatedly play and stop the material during a quiet passage. If we hear a click or a pop when starting or stopping, the DC offset should be filtered out. The best solution for DC offset is a very steep high-pass filter below 10-20 Hz. The jury is out on whether or not this should be a linear phase filter. Listen and decide.[2]

II. "Subjective" Techniques

Workflow: Analog, Digital or Hybrid?

Everyone has his own style of working. Mastering engineers may prefer to work with analog processing gear for many reasons, not the least of which are *comfort and speed*. When you work 8 hours a day, you cannot afford to have an uncomfortable or time-consuming workflow. Analog gear has knobs that are easy and quick to grab, and it produces instant results without crawling through menus or digging through frustrating multiple choices ("now where did they hide that option?"). Have you ever been in MIDI hell? Computer hell? But, to be fair, some outboard digital processors are as easy to use as analog, such as the Weiss Gambit series, with its touch-sensitive knobs, or the TC System 6000, with its ergonomic Icon touchscreen. It's not necessary to pour through a manual to figure out how to use them: they're intuitive and easy to use. Digital gear has convenience features, like memory recall, that most analog gear does not have. A growing category of outboard is hybrid gear, i.e., digitally-controlled analog circuitry. Hybrid is audio nirvana, the best of both worlds, assuming that the digitally-controlled circuitry sounds as good as pure analog. In my opinion, most hybrid gear does not sound as good as pure analog, but I have found some rare exceptions. What remains is a question of sonics. Digital emulators continue to get better, and the sonic differences smaller and smaller, but frankly for me some tube and solid-state gear are irreplaceable. No digital emulator has the fullness, warmth and depth of a Pendulum ES-8 or the grit of an API 2500. So I firmly believe that a complete mastering house needs to have a collection of digital and analog processors. It is possible to push audio levels of analog compressors with fewer artifacts than the same amount of compression in the digital emulator; with analog processors it's easier to get "attitude" without digititis, especially in rock and roll. On the other hand, I find that when you need transparency and delicacy without coloration, digital processors are often better choices than analog. That's why I recommend an adaptable hybrid workflow. In Chapter 22 I discuss the reasons why some digital emulators don't sound as good as the equivalent analog gear.

Monitoring: 16 bit or 24 bit? 96 kHz or 44.1 kHz?

In later chapters I will discuss wordlengths, dither, and sample rates in detail, but right now let's discuss the practical techniques. For clarity, I will abbreviate format names: "2496" means "24-bit/96 kHz." One issue in mastering is that we often work at a high resolution, e.g., 2496, while producing material that needs to be at a lower sample rate and wordlength. By working at the higher resolution we get greater purity of tone and soundfield depth. In one project we might have to produce end masters in all these formats:

- High Resolution masters (for DVD, HD Tracks, Pure Audio BluRay, etc.)
- LP masters (which are often 2496)
- Mastered for iTunes (which are either 2496 or 2444), to be converted to AAC format by Apple
- CD masters (1644), also used as sources for digital download services that will convert to AAC or mp3

In general, lower resolution masters exhibit a smaller soundstage depth and width. The 16-bit masters also have a different texture or sound character. However, it has proved useless to attempt to compensate for these losses using tricks (such as equalization or ambience retrieval) on the high resolution side. Boosting high frequencies does not work: it usually makes the 1644 sound harsh, and hurts the sound of the high-res. Ambience retrieval techniques also hurt the sound of the high-res if it already has an optimum amount of ambience and depth. The key is to know what kinds of losses can occur (there will always be some audible loss), and what things we can do in the succeeding steps to minimize those losses. For about a year my practice was to work in a 96k environment while auditioning a dithered 1644 through a Weiss hardware sample rate converter and a second instance of Sequoia. This let me judge the losses and make a preliminary choice of 16-bit dither. But soon I learned that monitoring the lower res during production doesn't buy me any advantage. I could wait till the end stage to choose the dither, which would not affect any of my previous decisions. So today I produce and audition the best-sounding high-resolution master possible, then take the signal downward with the most transparent method. For example, I create the 2496 high-res master and take it down to 3244 with a superior-quality sample rate converter, such as the Weiss Saracon or Izotope's resampler. I then examine the levels for clipping using either Apple's **MFIT** (Mastered for iTunes) tools or the Fraunhofer Pro-codec plug-in available from Sonnox. Next, I decide whether to reduce the level to prevent clipping of subsequent AAC or mp3 files, and dither down to 2444 for MFIT. I then pick a 16-bit dither that helps the CD master sound as close as possible to the high res master, and cut a CD master. We will examine the fine details of wordlengths, sample rates, dither, clipping, and AAC in later chapters.

Achieving Dynamic Impact (Punch)

How do we create a *punchy master*? Here are some opinions:

> If you heard the unmastered mixes, you'd probably find them considerably punchier than the mastered version. — John Scrip

> Punch is... the right ratio of transients to well-timed compression. Too much of either and the punch is lost. EQ can clean up the punch already in a mix. — Brian Lucey

> Punch is first captured at the recording stage. If done right it is retained at the mixing stage. If done right again it is retained at the mastering stage... [Only] if it's in there can I enhance it.
> — Larry DeVivo

As we described in Chapter 6, when using dynamics processors, punch can be retained and sometimes improved when their attack and release times are optimized to permit both transients and sustained sounds while maintaining or enhancing the overall dynamic shape of the song. Does this mean that a punchy, snappy, dynamic master can't be produced from a MIDI'ed rhythm section based on 808 kick, synthesized sampled bass and sampled handclaps? Frankly, that's an uphill climb, it requires a team of engineers with ability and knowledge, and it wouldn't hurt to have at least one focused lead or rhythm line played on a standard instrument.

Less is more. I once produced a competitive alternative-rock album and at the client's request, I matched its sound to the competition, which was extra crispy and extra hot. I sent a reference to the client with this caveat: "Guys, as the album mastering progressed, the more I listened, the more fatigued my ears got. It's a combination of the strong compression and the slightly brighter-than-usual sound. So if you want to go for a sound that breathes more, and is a bit lower still, I'm definitely for it!" After they listened, they said, "Go for it, Bob!" The new mastering was objectively 2 dB lower in level, warmer, and with distinct bottom, but sounding so much better that nobody felt anything was missing. In fact, I got the opposite reaction. I told them, "It *really* kicks ass now! It has impact and punch and clarity, actually sounds more aggressive than the previous version because giving it room to breathe allows it to really kick. I'm loving it." After they heard version 2 their reply was very emotional: "Bob! Dude!! I'm only two songs deep and it sounds *huge*!" And "Speechless dude . . . awesomeness, you rock . . . " It's not only gratifying to get those kinds of compliments: it also shows that too much compression is self-defeating. If you want a recording to sound huge, remember: *less is more!*

Fattening With Tubes

As I mentioned, one good thing about analog processors is that they do not exhibit *digititis*, which I define as the inharmonic distortion caused by the artifacts of certain kinds of digital calculations (to be explained in Chapter 22). Some analog processors are transparent, and some change the sound in an enjoyable way. It's important to have different types of processors for different styles of music, and also to avoid processing when the music demands complete transparency and no coloration. Everyone has his favorite tube processor. For subtle color, the versatile Pendulum OCL-2 is a *near-transparent* tube circuit. For transformative work, the Pendulum ES-8 is a colored vari-mu circuit with input transformers that can add a silky, liquid, creamy quality to the sound and a lush bottom by virtue of its unique transformer distortion. Even when the compression is turned off, passing signal through the ES-8 is transformative. But still, it's useful for a lot of material as the ES-8 does not sound too "vintage." The transformers' phase shift shuffles the image in a pleasant way, a bit of pleasant low frequency distortion that fattens the sound as well as reduces the peak to loudness ratio about 2 dB without even using the compression. I often run a recording through the ES-8 without its compression and let the transformers and tubes perform the transformation. Both of these tools can also help deal with harsh digital recordings: in some cases, miraculous results have been achieved.

Sample Rate Converters in a Mastering Chain

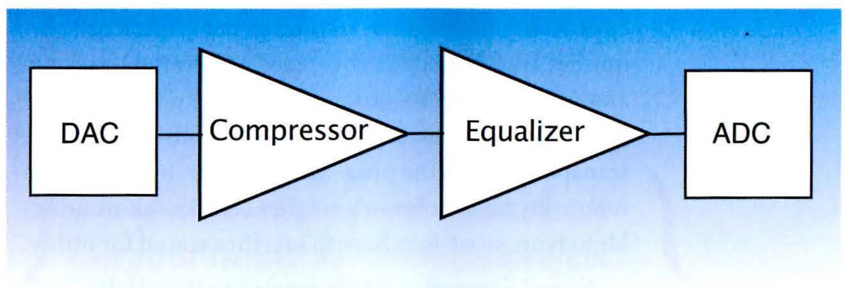
In Chapter 3, we briefly discussed upsampling in the mastering session to improve purity of tone. Here we examine the practical tradeoffs. Many processors, for example the Weiss, internally double sample when fed single sample rates. This means that if we digitally process a 44.1 kHz source, pass it through three Weiss processors, then back to the DAW, it has been upsampled and downsampled three times in a row. The Weiss uses extremely high-resolution processing and it is very difficult to detect degradation if any, but I still advise that you first upsample the material to avoid unnecessary DSP calculations. Though the differences may be subtle, in a cumulative chain, less DSP can sound more transparent and it avoids creating a colder sound. Instead of upsampling, some mastering engineers perform analog processing first, then use an ADC running at a higher rate. This effectively upsamples prior to any digital processing.

Patching Order of Processes

Sometimes it's better to compress before equalizing. For example, if you're using the EQ to enhance the level of an instrument or to add pressence, a compressor after the EQ might undo the effect of the equalizer by pushing the strongest sound downward. Sometimes that is the desired effect, but equalizing in front of the compressor is usually not a problem unless some emphasized frequency range causes the compressor to overreact. I almost always put sibilance controllers early in the chain, so they will operate with a constant threshold (sensitivity) regardless of how other devices are adjusted. Parallel compression can work nearly anywhere in a chain before the limiter, but I usually place it early in the chain. The analog processing portion of the chain (see figure below) can go nearly anywhere, but the digital peak limiter must be last. If you've placed a downsampler after the peak limiter, be sure to use an upsampled brickwall limiter that takes into account the level rise from later filtering processes.

Pitch and Time Correction

It is impossible to fix the relative pitch of a vocalist once he's mixed with other instruments, so mastering engineers are not often called upon to correct pitch. However, in isolated *a capella* moments, or when an



D/A/D Processing Chain. The choice of whether to place compressor or equalizer first depends on the intent. See text.

entire section of a tune is off-key due to an edit, we can make corrections in mastering. Pitch and time correctors (e.g., Autotune, Melodyne) are now quite sophisticated, and we can successfully use one for short periods; however, no solution is transparent, and some degradation can be heard in a high-resolution environment.

Pitch and speed correction altogether. Many engineers forget that the easiest and most transparent method is to change *both* the speed and pitch at once, like playing an analog tape recorder faster or slower, avoiding the severe manipulation of a repitching device. It's usually acceptable to the artist as well. Changing the speed also changes the sample rate, so we perform an asynchronous sample rate conversion, then reinsert the material of the "wrong" sample rate into the EDL. This technique can sound much better than a pitch corrector if a good SRC is used, thus avoiding the glitchy sound from splicing that all pitch correctors use. Many DAWs have available an asynchronous SRC, which allows minutely changing sample rates. A well-made software ASRC can sound as good as a synchronous SRC (the latter only permits integer multiples). However, current chip-based ASRCs are comparatively resolution-challenged.

Pitch correction while retaining the speed. This approach is not as good-sounding, but is often required. If pitch needs to be raised or lowered the same amount for a long section, I nominate the TC System 6000 VP engine as the best choice, because it's more transparent than the plug-ins. However, it cannot ergonomically raise or lower a single note like Autotune or Melodyne, so we may have to sacrifice sound for utility.

Speed correction while retaining the pitch. This process, called **time correction,** is probably the worst-sounding. I have not heard a DAW-integrated time-corrector that's as transparent as a two-step process. The first step is to alter the speed and pitch together using a resampler (sample rate converter), then to repitch using a good external pitch changer like the TC System 6000.

Bottom line: When it comes to pitch, it's better to let sleeping dogs lie.

III. "Remixing" at the Mastering Session

Introduction

Everyone has heard the expression "we'll fix it in the mix" — It's not the right solution for a bad performance, but it happens all the time. Similarly, leaving a problem for the mastering engineer to "fix" is not a good idea, but this happens, too — sometimes because of lack of time, and sometimes because the problem was not perceived during the mix. Given the challenge, we'll find a way to improve, even fix, sonic problems. Here are some approaches.

Vocal Up and Vocal Down Mixes

The lead vocal is considered "king" in nearly every style of music. That doesn't mean the vocal has to sit on top of everything else, but the dynamics of the vocal should contribute to the drive of the song along with the rhythm: the vocal should not be pulled along by the instruments. My principle is that a vocalist should neither be so low that she's struggling to be heard, nor so loud as to diminish the impact of the band. This works for most styles of music. A source mix can have a perfect lead vocal level, but occasionally after mastering processing, it might come up or down relative to the instruments, or the producers might change their minds after reflection. This is why we always recommend that the mix engineer produce vocal up and vocal down mixes, typically by 1/2 dB, as safeties. If more than 1/2 dB is needed, then probably not enough attention was paid to the vocal level in the first place. Keep in mind that the vocal level standard varies with some music styles. Some styles focus the vocal so loud that it diminishes the band to the point that it is no longer driving the impact; I've found this phenomenon in U.S. pop-Country music, occasional pop- (diva-driven) R&B, and in Greek popular music.

During mastering, we may want to insert pieces of the vocal up or down mix. For these decisions, we always collaborate with the producer, who may have purposely chosen a vocal level for stylistic reasons or to deal with a pitchy soloist.

Mastering from Stems

Stems are a special kind of submix. For example, a *lead vocal stem* and an *instrumental stem*, which when summed will equal the full mix. The mix engineer should supply stereo stems, even for mono instruments, to avoid a potential 3 dB ambiguity, since the mastering engineer would not know the panpot law of the console used by the mixing engineer preparing a mono stem. Generally, stems are wet, with the vocal stem having its own reverb, and so on. Each stem should contain unique elements with no overlap: otherwise there would be potential for comb filtering or unintended balances. Stems should be sample-accurate and begin at the same timestamp, which disqualifies 2-track analog tape as a stem medium (except for the most adventurous). One way to produce stems is to make multiple passes of the mix, each time muting different elements. It's slow, but it does guarantee that the sum of the stems equals the full mix in level, tonality, and amount of reverb. An even better and faster way to make stems is to mix "film style" to dedicated stem tracks, which are assigned to the stereo bus for monitoring. In this way you can punch into the mix when you make a revision, without having to run the entire mix. Then you can refine the edit at the punch. At the end you can consolidate the edits, then high-speed bounce to consolidated stems.

When the mix is done, it should be done! Stems are generally not intended for remixing or to "fix" the mix. I try to avoid stems, because they prolong the decision-process, but I'm not shy to suggest stems when I hear a weak mix that the producer is not going to get better within his time limit. Clearly, this stretches the role of the mastering engineer. The last thing we want to do in mastering is second-guess what the mix is supposed to be, but rather allow for minor tweaks if the mastering processing changes the mix slightly or to help present the intended mix in the best light. When clients see what we can do in mastering, they are tempted to revisit the mix, but we should discourage them from opening that can of worms. Anything may be possible; but our job is to help them maintain perspective and try not to wear the mixing engineer's hat while mastering. For as soon as we start concentrating on "the snare drum is too loud," we begin to lose the mastering engineer's goal: how to best present the sound and feel of the existing mix. In fact, I can't effectively wear both mixing and mastering hats at once, and I run a separate mixing session if there are more than two or three stems.

> *"If it takes me more than three or four hours to master a project, there is something wrong with it."*
> — Glenn Meadows

Not every mastering engineer is comfortable working with stems, because not every mastering engineer has been a mixing engineer. But stems can let us perform mastering processing with less compromise. Consider a live recording that I once mastered: it had a very heavy high hat and ride cymbal due to acoustical issues at the gig. I asked for a vocal and instrumental stem, then processed the instrumental stem with a high frequency compressor. This left the vocal untouched. Another example: not everyone has the pristine monitoring environment to make proper bass level or EQ judgments, so I sometimes suggest sending a separate bass instrument stem when the level or EQ of the bass is not optimum. Leveling or EQ'ing the bass instrument can produce much better results than EQ'ing the whole mix, as long as we do not step on the producer's intent, but rather enable the producer's intent. It's all part of the collaboration: using experience, common sense, and communication between the mastering engineer and the producer.

Stems generally do not work if the mix was made with aggressive bus compression, since the bus compressor behaves differently when fed the full mix than when fed the individual elements. But if only light bus compression was used to "glue" the mix together, mastering compression can do an equivalent or even better job.

In addition to the full mix, two or three stems may be provided, each mixed in stereo (or surround) incorporating its own reverb:

- Lead Vocal (sometimes labeled *a capella*)
- Instrumental
- TV (instrumental plus chorus)

Sometimes, the instrumental may be split into rhythm and melody stems, or into three stems with the bass instrument separate. But additional stems should be considered only if the mix engineer has doubts about a particular instrument as it inevitably leads to some remixing at the mastering session. Usually the instrumental stem is not needed in mastering, as the sum of the lead and TV equals the full mix. However, if the client complains that the chorus is too loud, we could add in a pinch of the instrumental and drop the sum of instrumental plus TV. If a full mix and TV are provided, but no lead vocal, it is possible to increase the lead vocal by subtracting the TV from the full mix (add TV with inverted polarity) and then raising the overall level. Given a full mix and a vocal stem, we can add or subtract vocal. As soon as we get into these kinds of maneuvers, we must match up levels to get a "base plane" from which to work.

When Stems are Not Available

Whether we like it or not, we are often asked to find ways to fix a mix, to try to isolate, raise, or lower a particular instrument, when a remix is not possible. Probably the first instrument affected by the attack time of a mastering compressor is the snare drum, which allows us to adjust its level to some degree, especially if we compress (or expand) in the 1-2 kHz range. We can greatly improve the impact and clarity of the rhythm, particularly the snare, without changing the tonality of the vocal, by using upward expansion with a relatively short attack time. Or we can pull the snare back a hair if it's interfering with the vocal, using a compressor with a relatively short attack time. It's frequently possible to enhance or punch the bottom end of the bass drum without significantly affecting the bass instrument, by using a low-bass-frequency compressor with a relatively long attack time in conjunction with an equalizer. If the bass drum is too loud, as a supplement to EQ, try a narrow-band compressor centered around 60 Hz and stay below the bass instrument. If the bass instrument is too "jumpy" and loud at times, one band of a multiband compressor can do the job, as long as we don't need too much correction; 1 dB gain reduction goes a long way, but more than that can start to drain the life out of the bottom end of the mix. Such "fixes" concentrate on the low end of the bass and bass drum. However, if the bass needs presence or the bass drum needs more beater clarity, an EQ boost is bound to collide with the vocal, snare or keyboards and cause trouble.

MS Mastering

It's always better to ask for a remix, but when a remix is not possible, the MS technique can be used to gain a bit more control over the separate elements in a recording. MS stands for *Mid/Side*, or *Mono/Stereo*. In MS microphone technique, a cardioid, front-facing microphone is fed to the **M**, or mono channel, and a figure-8, side-facing microphone is fed to the **S**, or stereo channel. A simple decoder (just an audio mixer, also known as a *matrix*) combines these two channels to produce **L**(eft) and **R**(ight) outputs. Here's the decoder formula: M plus S equals L, M minus S equals R.[3] It's possible to create an MS encoder or decoder using faders in a mixer and selective polarity inversion, but why go through the trouble building a decoder when the MS revolution is now upon us: For example, the Weiss DS-1 Mk3 can perform independent single-band left-, right-, M- or S-channel compression or expansion. The TC Electronic System 6000 MD4 can perform multiband MS compression. The Weiss EQ-1 (hardware), the DMG Equilibrium plug-in, and the UAD Brainworx bx_digital plug-in can separately equalize left, right, M, or S in any band(s). Plug-ins from Waves and others enable separate M and S channel inserts in the plug-in chain. Analog mastering consoles from Crookwood, Dangerous Music, Manley, and Maselec implement MS inserts, enabling MS compression with outboard analog processors.

Careful, conservative use of MS tools can turn a good recording into a great one, or save a so-so recording from the dust-heap. Keep in mind that we must approach any "remix" task with experienced ears and respect for the producer's wishes. Examples:

- Width control: all stereo width controls are MS processors in disguise. Any processor that separates signal into M and S components can affect width: more S level increases the width, and vice versa.
- Boomy bass: a client once mixed in a bass-light room, and his bass was very boomy, up to about 180 Hz. At first the vocal level came down a bit by correcting the boomy bass through EQ alone, but I was able to produce a well-balanced master with little compromise to the other elements by combining M and S-Channel selective bass frequency cut with M-channel selective mid-frequency boost.
- Weak vocal level: Faced with a light, center-located vocalist, we can raise the M channel. The vocal level comes up, as does the bass (usually) and every other centered instrument. Keep in mind that raising any center-located instrument via the M channel reduces stereo separation and depth and makes the program sound more monophonic. However, if we restrict the boost to the midband, e.g., boost the lower midrange and/or the presence range in just the M channel, this brings up the center vocal with less effect on the other instruments, and less loss of stereo separation.

• Lead singer vs. background singers: It's possible to alter the balance between center-located lead vocalist and side-located background singers, even varying the MS ratio between verse and chorus of the song to bring out the chorus when needed. This idea may sound scary to a savvy producer, but attractive to one who didn't produce the perfect mix.
• Splashy crash cymbals: Instead of using overall high frequency equalization, equalize the S channel, provided that the crash cymbals have been mixed towards the sides of the stereo image (which is often the case).
• Spread the cymbals without losing the focus of the snare.
• Tighten the bass image without losing stereo separation of other instruments.

MS Compression. Consider a mix that sounds great, but the lead vocal is slightly buried when the instruments get loud. By using MS compression or equalized MS compression (a multiband compressor in MS mode), we can isolate compression to the S channel or to an equalized portion of the S channel. This delicately brings down the sides (effectively bringing up the center) only when the signal gets loud.[4] Or, if the vocalist overpowers the band on the peaks, we can compress the M channel and/or expand the S.[5] Equalized MS compression (multiband technique) keeps the bass and treble ranges from being affected by our vocal (midrange) compression. In other instances, we might achieve a better kick drum sound by compressing only the low frequencies of only the M channel. S-channel parallel compression can be used to enhance and increase the spread of low level ambience, which can sound majestic. The possibilities are solely limited by our imaginations.

MS tradeoffs. Listen carefully for the MS tradeoffs with experienced, trained ears! When raising the M channel to increase the vocal, the stereo width narrows, and vice versa. Changing the width alters the mix. Mix engineers who rely largely on nearfield monitors often produce stereo-compromised mixes because listening on nearfields is like having a big set of headphones, which exaggerate the stereo separation. Using MS to increase the width is initially attractive, but can easily produce an unfocused, phasey, vague stereo image. Therefore more than about 1 dB of MS variation is usually a bad idea. By comparison, if the fixes had been done during the original mix session, there would be no tradeoffs. Only instruments which are perfectly centered or side-located are good candidates for MS correction. The K-Stereo processor, which I developed (full disclosure), is a good companion or substitute to MS processing, because it can help the depth, width and apparent stereo separation of a mix without seriously affecting the mix levels. Or, compensate for the depth and separation losses when it was necessary to raise the M level.

Automating the MS correction. MS variation can be accomplished by automating a plug-in such as the Waves S1, or directly in an EDL without using any processor. One way to raise a (centered) vocal is to add a duplicate of the material in another stream, with the channels reversed. Add this in at as low a level as possible (typically -12 to -16 dB), for if taken to an extreme it will turn the entire material to monophonic. Nowadays many mastering DAWs make automated MS correction easy by incorporating width controls, which are easily automated. Sequoia's object-based processing lets us make a cut in an object, change the width or other processing just for the duration of that cut, and crossfade between the altered object and the surrounding objects. For the object that needs a raised vocal, we may add a pinch of a stereo ambience processor to compensate for loss of ambience, and lower the bass gain to reduce center-channel bass build-up, just for the duration of the object, until the transitions are invisible to the listener.

MS with reverb. Clients sometimes ask us to enhance the vocal reverb without performing a remix. Consider a recording with vocal in the middle and a guitar on each side. A mono-in/stereo out reverb will automatically exaggerate the M-channel content.

IV. Making It Louder with the Least Compromise

The Red Light District

In Chapters 17-19 we discuss the loudness revolution. Technologies are rapidly moving into place which normalize the loudness on the media, so clients in the immediate future will not be asking for a louder master. But currently the most frequent question they ask is "how do you make a master sound louder?" My most frequent answer: "by turning up your own volume control!" This question is motivated more by fear and doubt than by the facts. Most good mastering engineers bring a program up to its loudness potential with all of our skills, which means that if the level is raised even one dB more, the sound quality will deteriorate. At home and on the media, the results of trying to make a master louder usually do more harm than good: On terrestrial and satellite radio, overcompressed music sounds distorted due to the processing (see Appendix II). On loudness-normalized media and Internet radio, it sounds smaller and less clear, but moderately-compressed competition sounds big, clear and impressive.

I try to advise clients on the situation, but if in a competitive environment the client requests an even louder master, then we can try to raise it with the least sonic compromise. So: caveat engineer—apply the following techniques with caution, and be aware of their limitations.

"When raising the level, listen, don't look at the loudness meter."

Simple Equalization

The Fletcher-Munson effect (revised in ISO 226:2003) dictates that high frequency energy produces more loudness than low frequencies at the same SPL. The first thing to try is a high pass filter, then a high frequency or presence boost if necessary. However, both of these can produce the opposite effect when played on FM radio with its high-frequency pre-emphasis. An overly bright, zippy song that sounds loud in the mastering room will get reduced by processing when it reaches FM radio. If the high frequencies sound fatiguing in the control room after multiple plays, it's a sure sign the song will distort on FM radio.

Parallel Compression

This is probably the single most potent technique to add loudness and power to a master. If performed on material which only needs moderate compression, parallel compression with the right time constants can have a singularly positive effect, creating not just louder sound, but also esthetically better sound with more *body* and *punch*. Since it operates first on the lower levels, we can often obtain 2 or 3 dB of additional loudness with little squashing of the higher levels. In a metal or rock piece, adding a small dose of parallel compression can supply additional strength, in conjunction with the right kind and amount of downward compression.

Harmonic Synthesizers

Harmonic synthesizers can supply edge, depth and a form of compression, unlike a simple equalizer. The down side is that they can easily produce a harsh, thin, edgy sound (See Chapter 22).

Digital Peak Limiting

Keep in mind that if every mastering engineer is already using peak limiting: every runner in the race has already taken steroids, so there is no performance advantage! 1 or perhaps 2 dB of limiting may get the program level up without taking the sound quality downhill; more than that can easily produce wimpy, unclear, less effective sound quality. Carefully compare with less limiting and you may be surprised. When raising the level, listen, don't look at the loudness meter. Does the raised version actually sound louder, or just measure louder?

Clipping

Clipping can add an edge, increasing apparent loudness and definition, but it can also have the opposite effect! Since clipping is a form of limiting without a defined attack or release time, it can be a lesser evil, with less *clamping effect* than a peak limiter. But like limiting, it can rob a recording of impact and important microdynamics. At double sample rates, *digital clipping* produces less midband distortion and artifacts than at single sample rates. Be aware though, that what sounds "loud" due to clipping in the control room will sound very harsh once it hits a codec (e.g. AAC) or FM radio station processor: you can't violate the laws of physics (See Chapter 16). Audition clipping through the codec.

Genuine *analog-domain clipping* is the mastering engineer's secret weapon, because it does not exhibit any digititis. Digititis (inharmonic distortion) is caused by the beating of naturally-occurring harmonics against the sample rate, which occurs in typical digital compressors and limiters. Oversampling helps reduce digititis as we will learn in Chapter 22. However, when clipping in the analog domain, the extra harmonics do not beat against the sample rate, so they do not cause any inharmonic distortion. When the extra harmonics are fed into the ADC, they will either be filtered out or left unaltered.

Depending on the complexity of the musical material, clipping can be masked, or it can be revealed. In other words, 1 dB of clipping probably sounds horrid on an acoustic piano solo, but may be completely masked when the clipped piano is mixed into complex music. To my experience, run an ADC at double or higher sample rate for the fewest clipping artifacts. If we receive a clean mix, we *may* get away with two stages of clipping: 0.5 dB of oversampled digital clipping and up to 1 dB of analog clipping. However, this combination can induce serious distortion and so the sonic quality deteriorates extremely fast. Analog domain clipping is most effective if there is no further processing, or if the client insists on still more level, in front of a digital peak limiter with (hopefully) no more than 1 dB of action. As always, listen on high-resolution, high headroom monitoring to confirm the clipping/peak limiting is not making the sound distinctively worse. Make your comparisons carefully, because you can be surprised to learn that taking away the limiting or the clipping sounds louder if you have already exceeded the program's loudness potential.

To ensure the ADC has a controlled amount of clipping, first calibrate the chain so that the DAC operates into the ADC at unity gain, with all processors bypassed. But equalizer boosts can easily push an ADC too far. So it is difficult to assess objectively when the ADC ceases to be "a little oversaturated friend" and becomes our "seriously clipped" enemy. A digital peak meter on the ADC's output tells us nothing since the ADC's maximum peak level is 0 dBFS regardless of the input. We need a dependable analog-domain peak meter that clearly tells when the analog signal has exceeded the amount which would clip the ADC and by how much. Currently I use an indirect method by inserting analog attenuation in front of the ADC until the digital peak level goes below 0 dBFS. This lets me determine how much I have been pushing the ADC without the attenuator.

Chapter 16 will illuminate the technical reasons some kinds of clipping are completely invisible to the ear, while other kinds of clipping sound distorted.

Don't Spend it All in One Place

We can produce the loudest-sounding master with the fewest compromises by using a small amount of several processes in a row, rather than engaging just a single process. In all cases, the more *open* and dynamic the mix is to begin with, the easier it is for us to produce a "loud" master. A mix that has already been clipped or limited produces a smaller-sounding, harsh, and softer master. That's a hard lesson to communicate to mix engineers who emulate distorted masters and then send them off for mastering. Please don't do that.

1
The jury is out on whether the ear truly can detect absolute polarity. We surmise that it is the better loudspeaker that reveals polarity differences, but it may be because the loudspeaker is defective or non-linear. A displaced or non-centered voice coil could react differently to positive- versus negative-going signals. So, before claiming that your loudspeaker system reveals differences in absolute polarity, perform tests for distortion with asymmetrical test signals.

2
DAWs with a DC offset correction option attempt to fix it by centering the average center of the waveform on the 0 line. This is ineffective, because ADCs tend to drift, so there will still be some remaining offset. So I recommend the high-pass method when there is severe DC offset.

3
To avoid overload, the safest formulas for MS are:

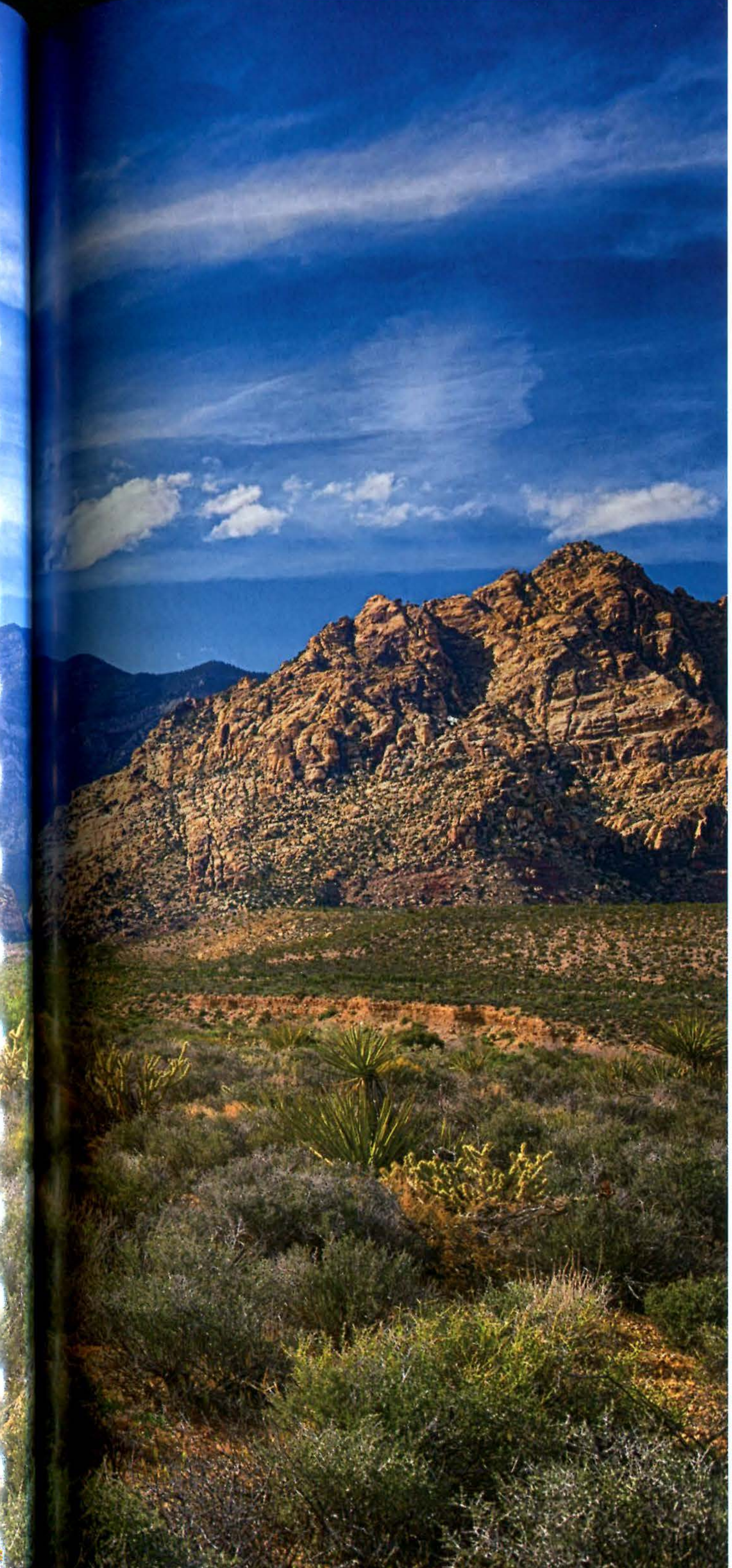
Encode: M = 0.5 * (L+R), which is 6 dB less than the mono sum. The encoder sums and attenuates by 6 dB. S = 0.5 * (L-R), which is 6 dB less than the mono difference. The encoder takes the difference and attenuates by 6 dB.

Decode: L = M+S. R = MS. Be aware that in this style of encode/decode, the two coders are not identical. However, in a floating point system, overload is not a concern: the two coders can be identical, 100% symmetrical, and even bit-transparent.

4
Remember that a downward compressor brings sound down when it goes over the threshold, so the loudness increase of the compressor is done by raising the gain makeup control, raising average levels but lowering the loudest. In the MS case, just 0.5 dB compression may be all you need to control that "lost" vocalist above the band.

5
If a unit that allows downward compression of M and upward expansion of S is not available, I may compress the M channel in one unit, then upwardly expand both channels in another. When properly adjusted, the net result is the same.

"Learning from your mistakes gives you room to make even bigger ones!"
— Murphy's law of experience



CHAPTER 10

How to Achieve Depth and Dimension in Recording Mixing and Mastering

I. Introduction

Some surround mix engineers are repeating their mistakes from two-channel work — panpotting mono instruments to discrete locations, then adding multiple layers of uncorrelated stereophonic reverb "wash," thinking it will create space and depth. It's important to learn how to manipulate the surprising depth available from the 2-channel canvas before moving on to multi-channel surround. Let's start by looking at some basic acoustic principles.

How True Stereo Recording Yields Better Reproduction

There's a big sonic difference between simple panpotted mono mixes versus genuine stereo recordings. Genuine stereo obtains a real sense of depth by utilizing the natural room acoustics and reflections from nearby walls. When all the musicians are playing at once, the natural interaction of microphones and acoustics (often called "leakage") helps to make a bigger sounding recording. Without this acoustical element, recordings tend to produce a vague, undefined image: the musical instruments and their positions in the soundstage are obscured and unclear. The ear's "decoder" craves delay information: a few well-placed echoes solidify and clarify the location of the direct sound, and help to distinguish one instrument from another in a complex mix. This is why a panpotted mono instrument (either close-miked, or one without stereophonic early reflections) is hard to locate precisely between two loudspeakers. Its location becomes ambiguous as listeners move away from the center seat (the "sweet spot"). But when the instrument and its surrounding acoustics have been captured in true two-channel stereo, the sweet spot widens and the instrument's location becomes more precise.

Engineers need to think beyond the panpot. Many pop engineers know it is possible to simulate depth artificially, using delays or artificial reflections, which help localize instruments to sound similar to "the real thing." We need to understand the principles of the Haas[1] effect, particularly when implemented binaurally. Most of this knowledge is best applied during recording and mixing, but we can also use it during the mastering session.

II. Early Reflections and Masking

Early Reflections versus Reverberation

At first thought, it seems that depth in a recording can be achieved simply by increasing the proportion of reverberant-to-direct sound. But the artificial simulation of depth is a much more complex process. *Early reflections* consist of the part of the room sound within approximately the first 5 (usually 15) to 100 milliseconds of the direct sound, meaning that the source and its reflections are highly correlated. Think of the early reflections as being *attached* to the direct sound. In a large and diffuse room, after about 100 milliseconds, enough wall bounces have occurred to create random (uncorrelated) *reverberation*. We can say this is *detached* from the direct sound. That is why the early reflections affect our perception of the depth and direction of the sound, giving it shape and dimension, while the reverb simply defines size of the space.

Masking Principle

The direct sound from an instrument can mask the direct sound of another; the direct sounds of instruments can also mask the reverberation of the room. *There are three types of masking:* **amplitude, directional, and temporal. Amplitude masking** occurs when a louder sound masks a softer one, especially if the two sounds lie in the same frequency range. This is why mixing engineers use equalization and filtering as well as level to separate elements of a mix. If the two sounds happen to be the direct sound from a musical instrument and the reverberation from that same instrument, then the initial reverberation can appear to be covered by the direct sound. When the direct sound ceases, the reverberant hangover is finally perceived.

Mixing engineers can add a small pre-delay between the direct sound and the onset of reverberation – a *temporal unmasking technique* that helps the ear to separate one from the other. A good *uncluttered* musical arrangement has built-in temporal unmasking, separating the timing and rhythm of the instruments, and making the mixing process easier.

Directional Masking during Stereo-to-Mono Reduction. In concert halls, to some extent in stereophonic reproduction, and to a much greater extent in surround-sound playback, our two ears sense reverberation coming diffusely from all around us, and the direct sound as having a distinct single location; there is little **directional masking**. However, in a monophonic recording, the reverberation is reproduced from the same source speaker as the direct sound, and so as the two sounds overlap directionally, we may perceive the room as being less reverberant. A very live recording hall is bad for mono recording, because reverberation will directionally mask direct sound. This is one explanation for the incompatibility of many stereophonic recordings with monophonic reproduction.

In 2-channel and multichannel recordings we can overcome directional masking problems by spreading artificial reverberation spatially away from the direct source, achieving a recording that is both clear (intelligible) and warm at the same time. One of the first tricks that mix engineers learn is to put reverberation in the opposite channel from the source. But though this can help to unmask the direct sound, it can produce an unnatural effect.[2] More sophisticated techniques use multiple delays or stereophonic early reflections to yield a more cohesive, natural result than is possible by simply using opposite-channel panning. Cheap reverbs containing no early reflections and a basic reverb "wash" muddy up the sound and deteriorate the depth. When you are looking to achieve natural depth and space as opposed to a special effect, it's better to have no reverb than a cheap reverb. Still, skilled mix engineers vary their colors, send different groups of sounds through different kinds of effects (some of which can be "cheap" effects), which helps to dimensionalize the presentation. True stereo reverbs produce better depth than mono-in/stereo-out models, although the mono-in/stereo-out TC VSS3 and the EMT 250 produce such great stereo ambience that they yield excellent depth only rivaled by the best true stereo or true surround models.

In natural environments, the early correlated room reflections are captured with their correct placement; they support the original sound, help us determine the distance of the sound source, and do not interfere with directional localization. The later uncorrelated reverberation naturally contributes to the perception of distance, but because it is uncorrelated with the original source, it does not help us locate the original source in space. The better the original room and the miking techniques, the more convincing the sense of space and the less artificial reverberation will be needed in post-production.

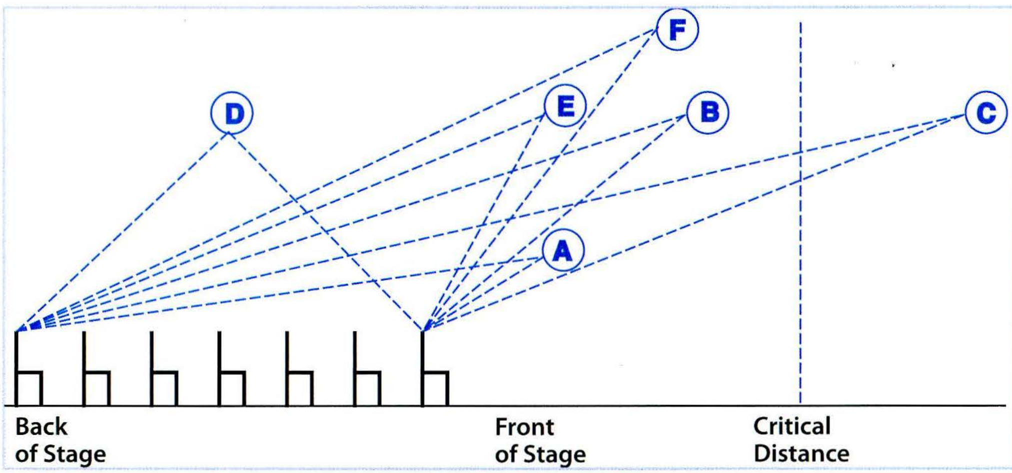
"It's better not to have any reverb than a cheap one!"

III. Recording Techniques for Depth

Recording in Natural Rooms

Balancing the Orchestra with only a few microphones (minimalist). The loudness of an instrument affects its balance in the mix; softer instruments also sound a bit farther away. But the primary influence on perception of depth and distance is the amount of early reflections and reverberation as well as the loss of high frequencies at a distance. A musical group is shown in a hall cross-section (pictured below). Various microphone positions are indicated by letters **A-F.**

Microphone position **A** is located very close to the front of the orchestra. As a result, the ratio of **A**'s distance from the back compared to the front is very large. Consequently, the front of the orchestra will be much louder in comparison to the rear, and the amount of early reflections reaching the microphone from the rear will be far greater than from the front. Front-to-back balance will be exaggerated. However, there is much to be said in favor of mike position **A**, since the conductor usually stands there, he purposely places the softer instruments (strings) in the front, and the louder (brass and percussion) in the back. Also, the radiation characteristics of the horns of trumpets and trombones help them to overcome distance, and the focus of the horn increases direct-to-reflected ratio. We take these factors into account when arranging an ensemble for recording.



Hall cross-section

The farther back we move in the hall, the smaller the ratio of back-to-front distance, and the front instruments have less advantage over the rear. At position **B**, the brass and percussion are only two times the distance from the mikes as the strings. This (according to theory) makes the back of the orchestra 6 dB down compared to the front, but in reality there is much less difference, because level changes less with distance in a reverberant hall.

For example, in position **C**, the microphones are beyond the critical distance — the point where direct and reverberant sound are equal. If the front of the orchestra seems too loud at **B**, position **C** will not solve the problem; it will have similar front-to-back balance but be more buried in reverberation.

Using Microphone Height To Control Depth And Reverberation

Changing the microphone's height allows us to alter the front-to-back perspective independently of reverberation. Position **D** has no front-to-back depth, since the mikes are directly over the center of the orchestra. Position **E** is the same horizontal distance from the orchestra as **A**, but being much higher, the relative back-to-front ratio is much less. At **E** we may find the ideal depth perspective and a good level balance between the front and rear instruments. If even less front-to-back depth is desired, then **F** may be the solution, although with more overall reverberation and a greater distance.

Directivity Of Musical Instruments

Frequently, the higher the mike is located, the more high frequencies it will capture, especially from the strings. This is because the high frequencies of many instruments (particularly violins and violas) radiate upward as well as forward. The ear perceives a brighter sound as closer, overcoming the distance. When the mike moves past the critical distance, we may not hear significant changes in high frequency response when height is changed.

The recording engineer listens and makes changes in mike placement based on these factors. The difference between a B+ recording and an A+ recording can be a matter of mere centimeters.

Mike Spacing, Pattern and the Depth Picture

Coincident Microphones. The various simple miking techniques we've been discussing reveal depth to greater or lesser degree. Microphone patterns which have out-of-phase lobes (e.g. hypercardioid and figure-8) can produce a holographic depth quality when used in properly angled pairs. Even coincident figure-8s provide as much or more of a depth picture than spaced omnis. But coincident miking reduces time ambiguity between left and right channels, and sometimes we seek that very ambiguity. With any given mike pattern, the farther apart the microphones of a pair, the wider the stereo image of the ensemble and the greater the hole in the middle. Instruments near the sides tend to pull more left or right, center instruments tend to get wider, more diffuse, and harder to locate or focus.

The technical reasons for this are tied in to the Haas effect (to be explained) for delays of under approximately 5 ms., vs. significantly longer delays. Very short delays between two spatially located sources produce ambiguous image location.

Spaced microphones. I have found that increased intermike spacing increases the center depth; for example, the front line of a chorus no longer seems straight: instead, it appears on an arc, bowing away from the listener in the middle. If soloists are placed at the left and right sides of this chorus, a rather pleasant and workable artificial depth effect will occur. Adding a third omnidirectional mike in the center of two other omnis can stabilize the center image and reduce center depth.

Beyond Minimalist Recording

Even after obtaining perfect balance, the engineer/producer often desires additional warmth, ambience, or distance. With time, he might discover a slightly more distant mike position with good balance, but time is precious during orchestral recording, and we hesitate to fix what isn't broken. Another call for increased ambience is when the hall is a bit dry. The engineer may try changing the microphone pattern(s) to less directional (e.g. omni or figure-8) but this then also requires a different spacing and angle. One solution is to use a Soundfield (tetrahedral) microphone and capture B-Format, whose direction, angle, and pattern can be adjusted in post-production.

Another solution is to add ambience mikes, being careful to avoid acoustic phase cancellation, which does not occur when the extra mikes are placed far enough to be in the uncorrelated reverberant field, or by applying the **3 to 1 rule.**[3] When these *uncorrelated* mikes are mixed into the program, direct frequency response should not deteriorate: we should simply hear an added warmth and increased reverberation.

Multiple Miking. While multiple close mikes destroy the depth picture, soloists do need to be heard, and for many reasons they are not always positioned in front of the group. When the soloist cannot be moved, plays too softly, or when hall acoustics make him sound too far back, then one or more supplemental *spot mikes* must be added. The depth image may seem more natural when the spot is a stereo pair, rather than a mono solo mike.

Apply the 3 to 1 rule, and listen closely for frequency response problems when the close mike is mixed in. This will (not surprisingly) appear to bring the solo instrument closer to the listener. If the level of the spot mike is not overdone, the effect is not a problem, as long as musical balance is maintained and the close mike levels are not changed during the performance. When mixing a recording made in a live acoustic space with supplemental spot microphones, try to match the panning of the spots to the virtual position heard by the main (distant) mikes (unless this produces an awkward balance). Since each spot mike also picks up ambience, the sum and spread of these accurate panning positions will enhance the stereophonic depth picture. In addition, leakage will end up in its proper virtual position. For example, if there is drum leakage in the piano mikes, the drum imaging and depth will be more accurate when the piano mikes are properly panned, producing less "smearing" of the drum image.

Delay Mixing. Adding a delay to each close mike to synchronize it with the main pair pulls the soloist back and helps to maintain natural depth, but we still need to hear some early reflections around the soloist, which hopefully arrive at the main pair with some strength. Otherwise, we should try a bit of artificial early reflections. To adjust the delay of the spot mike(s), start with a delay calculated by the relative distance between the solo mike and the main mike, then focus the delay up and down in 1 ms increments (even less if there is a comb filtering problem) until the sound is most coherent and focused, clarifying the sound of the soloist. It pays to record a series of clicks or hand claps before the performance and adjust the delays afterward until the sound is most focused. It also pays to record a person speaking or singing at each of the close mike positions and adjust the delays for minimum comb filtering and best focus between the close mikes and the distant mikes.

Dead Studios

Minimalist miking techniques do not work well in a dead studio. In a dead room, simple miking has no advantage over multiple miking with panpots because there are no early reflections. In this case, artificial ambience tools are required, and this is what separates the men from the boys.

IV. Adding Depth in Mixing and Mastering

The Haas Effect

The Haas effect can help increase definition, depth and fullness without causing masking problems. Haas said that very short echos (less than about 1 ms) produce an ambiguous (confused) image. However, echos from about 10 through approximately 40 milliseconds after the direct sound become *fused* with the direct sound. It also helps that the echos be diffused and a bit spread rather than specular or sharply-located. In other words, the ear continues to locate the original source at its original location, and the echoes only cause a loudness enhancement. This is already how the ear/brain works in a real room environment with its early wall and floor reflections. This is why the proscenium arch and side walls of a stage are so important to the musicians; these walls are near enough to provide sonic support yet far enough away to avoid comb filtering. Since the velocity of sound is approximately one foot per ms, 40 ms corresponds to a wall that's 20 feet (6.1 meters) distant and perpendicular to the direct sound.

Haas Delays in Mixing Enhance Spatial Qualities and Improve over Standard Equalization

In pop or classical mixing, we can use delays to take advantage of a very important corollary to the Haas effect, which says that fusion (and loudness enhancement) will occur even if the closely-timed echo comes from a *different direction*. The brain will continue to recognize (binaurally) *the location of the original sound as the proper direction of the source*. The Haas effect allows added delays to enhance and reinforce an original sound without confusing its directionality — just as long as the delay is not too long and the level of the delayed signal is not too loud. When the delay is too long or the delayed signal too loud, it starts to be perceived as a discrete echo; which we call the *Haas Breakdown* point. Long delays maximize the definition of the source, as long as we have not reached breakdown. The Haas breakdown point is shorter for percussive sounds; for example, sometimes only 15 ms is tolerable for a drum hit, while up to 30-50 ms is permissible for strings.

To take advantage of the ear's own decoding power during mixing, generally use panned and leveled delays in the 12 to 40 millisecond range. Haas delays are more effective than equalization at repairing the sound of a drumset that was recorded in a dead room. To create layers in the mix, put single delays on some instruments, and multiple (or no) delays on others; try doubler and quadrupler delay plug-ins with built-in panning, supplemented with the panpots in the console.

"A good artificial reverb is not just a sonic flavor, it's a powerful tool to help create depth."

Haas and mono-compatibility. When utilizing simple Haas delays, be sure to check the recording in mono for comb filtering. I tend to stay above 10 ms to improve mono compatibility. The more complex, diffuse and numerous the delays, the less likely that comb filtering will occur.

When mixing in surround, it is best to avoid power-panning (standard panpots) between front and surround: this produces a very ambiguous image, due to the extreme wide distance between front and surround speakers in the 5.1 format. As with stereo, *think beyond the panpot*. Virtual panpot positions do not image well: it is much better to use a surround early reflection generator, which produces a more stable image and allows a wider sweet spot. Early reflections avoid some of the comb-filtering which can occur when listening to a single Haas-delay in mono. In addition, early reflection-based positioning helps to locate objects between the loudspeakers for non-center-located listeners.

Artificial Reverb

Artificial Reverb: everyone uses it. There's "cheesy" reverb and then there's **REVERB!** A powerful, well-designed artificial reverb is not just a *flavor* to use, it's a *tool* that can truly enhance a recording. Mix engineers can use the computerized early-reflection simulations found in devices such as: the Bricasti M7, EMT model 250 and its UAD replica, Flux SPAT, and the TC Electronic model VSS4 and VSS6. These processors not only provide sonic flavor, but can also increase depth and soundstage size, and improve localization. Although the EMT is a mono in/stereo out device, it still adds a nice dimensionality, and to some extent enhances the localization of sources within the stereo recording due to the Haas effect of its early reflections.

Natural Equalization via Early Reflections

Think outside the box: I received a recording of a brass ensemble that the leader recorded in his living room surrounded by pillows and rugs (don't ask). Counter to intuition, pillows don't make the sound warmer: instead they make the brass sound brighter because it is dominated by the direct radiation from the horns. There were no early reflections to speak of in his recording, and, besides, if he had removed the pillows, the living room walls would have been so close as to cause comb filtering of the sound. Rolling off the highs (equalization) would not have cured the issue — the brass would still have sounded bright. Instead, I used the early-reflection generator in the TC Electronic VSS4, along with its excellent reverb. This completely transformed the recording, added depth and warmed up the sound in the same way that reflections coming from real walls would smooth, spread, and warm the sound.

Adding Haas Effect Processing During Mixing or Mastering

Here are four possible reasons why recordings we receive for mastering can lack depth, spatiality, localization and clarity:

1) The recording was made in a live room, but the mix engineer did not take advantage of the room's acoustics.
2) The recording was made in a dead room, and the mix engineer either used no artificial reverberation or cheesy reverberation that does not provide adequate early reflections or sense of depth.
3) The mix engineer monitored with nearfield loudspeakers, which fooled him into believing he had a wide soundstage (nearfield monitoring is like wearing big headphones).
4) The mix engineer did not take advantage of the full digital resolution of his work station and/or did not properly dither the outputs of his workstation (See Chapter 15).

In the first case, adding artificial reverberation or early reflections to a live room recording can muddy or defocus the sound if the mix engineer is not careful. If the room was good-sounding, then during mixing, carefully adding ambience mikes can work, but since most recordings are overdubbed in sections, it is not always possible to add room mikes as "glue" unless there was careful planning during each tracking session. An ambience extraction tool used subtly during mixing can extract and enhance the depth of the room. In the second case, we may suggest a remix with better reverberators. Or if the reverb used in the mix has a decent tonality but no sense of depth, we may try adding early reflections during mastering, using a superb reverb algorithm such as I mentioned above, provided

The Flux IRCAM SPAT. This is one of three feature-filled pages.

that adding the process to an already-mixed program does not dilute the mix. Early reflections added subtly can enhance the definition of a mix. The depth effect will be more convincing when the original room has some useful reflections, which we can combine with the artificial ones to enhance the reality. The VSS4 has a feature called *early decrease*. This prevents generating artificial early reflections that are already present in the source. All but the EMT have versatile control over the shape and amount of the early reflections. The Flux Spat provides separate control over the *cluster*, which is in between the early reflections and the later reverb.

Early Reflections to Enhance Localization

The TC VSS4 and VSS6 reverbs enhance localization in either stereo or surround by thinking beyond the panpot. The VSS4 provides several discrete positions for its early-reflection generator, letting the user precisely locate a source within the stereo or surround field. Our two ears do the "decoding," by detecting each wavefront and its location.

A remarkable tool to control localization is the IRCAM SPAT, a plug-in from Flux (above). In continuous development for over ten years at the French IRCAM institute, the versatile SPAT is multi-channel capable: It can be used to expand channel count and translate between different loudspeaker position standards. SPAT generates one of the most natural reverberation characteristics I have ever heard, and may be the best-sounding plug-in reverb ever designed. It goes beyond "just reverb" by panning and early-reflection-based localization. SPAT's early reflection generator is tied-in with its positioning algorithm. I have a bit of a quarrel with them over how their "running reverb" functions in conjunction with localization accuracy. It has to do with their definition of how "early" is "early," but let me summarize by saying that a SPAT user can easily lengthen the gap between direct sound and the first early reflection to avoid causing positional ambiguity or comb filtering, so if you know what you are doing, you can enable the power of SPAT's early reflections to enhance localization.

Using Frequency Response to Simulate Depth

In a natural acoustic environment, the apparent high-frequency response is reduced as the distance from a sound source increases. This provides another tool the recording engineer can use to simulate distance. An interesting experiment is to alter a treble control while playing back a good orchestral recording. Notice how the apparent front-to-back depth of the orchestra changes. We can use mikes with differing treble response, or during mixing, change the high frequency characteristics to move instruments forward or backward.

The Magic Surround

We can take advantage of the Haas effect to naturally and effectively convert an existing 2-channel recording to a surround medium. When remixing, simply place a discrete delay in the surround channels to enhance and extract the original ambience from a previously recorded source. No artificial reverberator is needed if there is sufficient reverberation in the original source. Here's how it works:

Haas fusion only works with *correlated* material. The ear fuses correlated sources with their delayed replicas (e.g., a snare drum hit) and so continues to perceive the direct sound as coming from the front speakers. But this does not apply to *uncorrelated* ambience — because the ear does not recognize the delay as a repeat, thus spreading, enhancing, and diffusing the ambience between the location of the original sound and the location of the delay. This was originally discovered by Madsen.[4] The wider the bandwidth of the surround system and the more diffuse its character, the more effective the psychoacoustic extraction of ambience to the surround speakers. Dolby laboratories called this effect *the magic surround*, for they discovered that natural reverberation was extracted to the rear speakers when a delay was applied to them. My patented K-Stereo and K-Surround processes start with and extend these principles.

Technological Impediments to Capturing Recorded Depth

Depth is the first thing to suffer when technology is incorrectly applied. Here is a summary of some of the technical practices that, when misused or accumulated, can contribute to a boringly flat, depthless recorded picture:

- Multimike techniques
- Small/dead recording studios or large rooms with

poor acoustics/missing early reflections

- Low-resolution recording media
- Excessive dynamic-range compression (which tends to amplify the mono information and bring everything forward)
- Overuse of "cheap" reverbs
- Using the same effect for all tracks
- Improper use of dithering, cumulative digital processing, and low-resolution digital processing (See Chapter 15)

Influence Of The Control Room Environment On Perceived Depth

At this point, many engineers may say, "I've never noticed depth in my control room!" The widespread practice of placing near-field monitors on the meter bridges of consoles kills almost all sense of reproduced depth. If you want to hear the depth in your recordings accurately, follow the advice in Chapter 21.

V. In Conclusion

Listening Examples

Here are some examples of stereo audiophile recordings I've made in great halls with minimalist miking that purposely take advantage of natural depth and space. Try to achieve this sound quality when working with artificial techniques. It helps to think in layers, "what's going to be in the foreground, the middle-ground, the background." Sara K. *Hobo*, Chesky JD155. Check out the percussion on track 3, "Brick House." Johnny Frigo, *Debut of a Legend*, Chesky JD119, especially the drums and the sax on track 9, "I Love Paris." Ana Caram, *The Other Side of Jobim*, Chesky JD73, particularly the percussion, cello and sax on "Correnteza." Carlos Heredia, *Gypsy Flamenco*, Chesky WO126. Listen to track 1 for the sound of the background singers and handclaps. Phil Woods, *Astor and Elis*, Chesky JD146, for the natural-sounding combination of intimacy and depth of the jazz ensemble.

To resurrect the missing depth in recording, mixing and mastering, use the highest resolution technology, best miking techniques, and room acoustics. Process dead tracks with Haas delays, early reflections, and specialized ambience recovery tools.

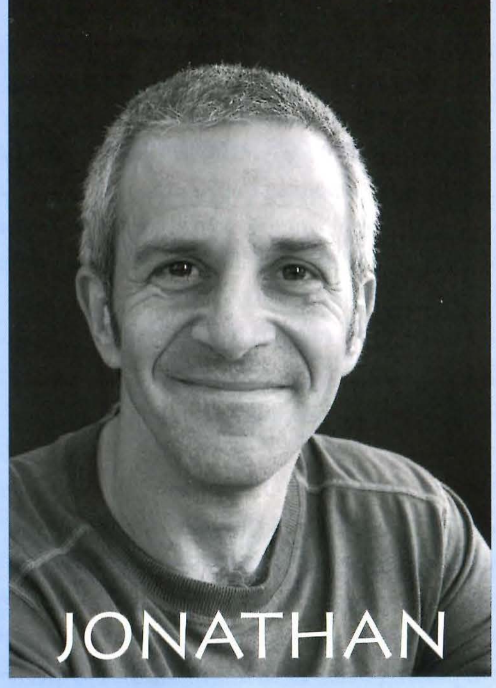
1
Haas, Helmut (1951), Acustica. The original article is in German. Various English-speaking authors have written their interpretations of Haas, which you can find in any decent textbook on audio recording techniques.

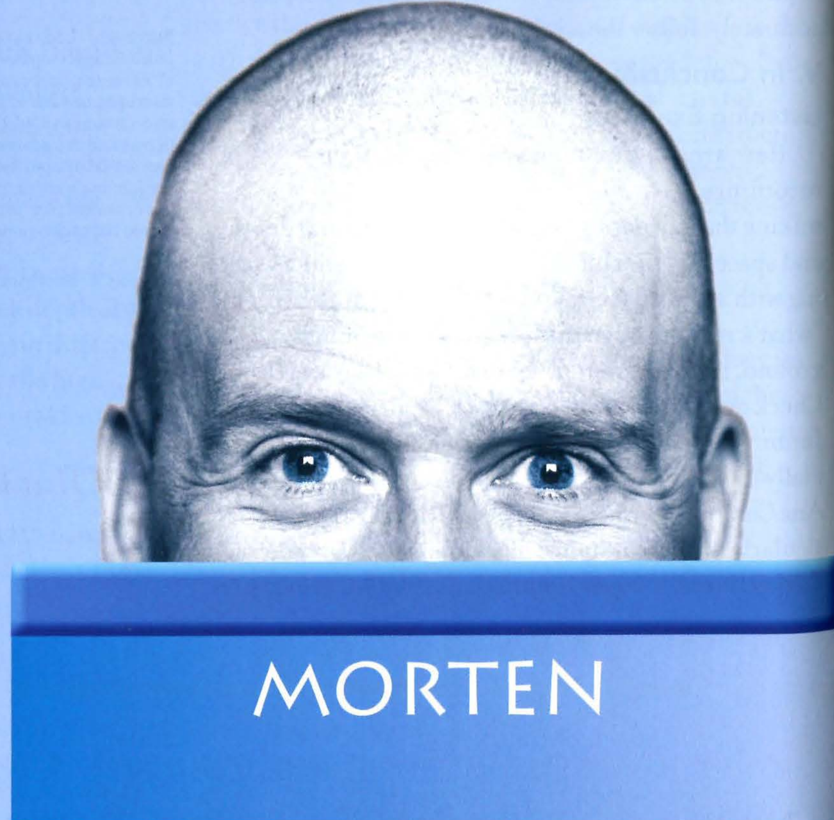
2
Even if unnatural, it can be interesting, nevertheless. Listen to 1960's-70's era rock recordings from the Beatles, Beach Boys, Lovin' Spoonful, The Supremes, Tommy James and the Shondells, where mono instruments or vocals are panned to one side, and often their reverb return completely to the other side.

3
Burroughs, Lou (1974), Microphones: Design and Application, Sagamore Publishing Company. (Out of Print). Burroughs quantified the effects of acoustic phase cancellation (comb filtering, interference) with real microphones and real rooms, and devised this rule: The distance between microphones should be three times the distance between each microphone and the source of the sound to which it is being applied. This is particularly important to avoid comb filtering when both microphones are feeding a single channel. When the microphones are feeding different channels (e.g. stereo), the degradation will be much less noticeable in stereo but still be a problem in mono.

4
Madsen, E. Roerback (1970, October), *Journal of the Audio Engineering Society*.

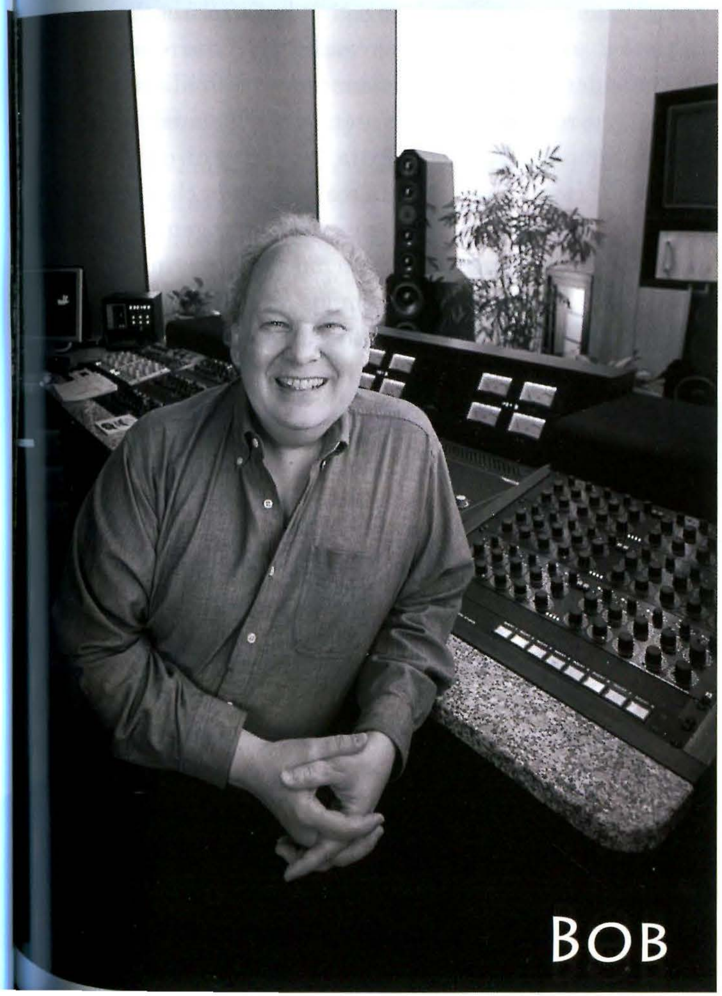
{ "Think Beyond The Panpot." }


RICH

JONATHAN

DAVE

MORTEN

CHAPTER 11

Surround Mastering: Q&A



Introduction

The concepts developed in the previous chapter help make both stereo and surround masters sound more dimensional and real. But surround sound by its nature is far more dimensional, realistic and much easier to make dynamic. Let's learn something about managing this exciting medium: In this chapter we meet five of the most talented and experienced surround sound engineers in the business. Each brings his own motif, but the themes are universal. Three are mastering engineers, one is a mix engineer, and one performs all three chores along with producing!

Dave Glasser Airshow, Boulder, Colorado
Morten Lindberg 2L Records, Oslo, Norway
Bob Ludwig Gateway Mastering & DVD, Portland, Maine
Rich Tozzoli Independent producer/mixer
Jonathan Wyner M-Works, Cambridge, Mass.

Let me start by pointing out that the 5.1 medium is a compromise between localization and envelopment. There are simply not enough loudspeakers to achieve both equally well, and thus the variance in opinion below about where to position the surround speakers. Jim Johnston says that 7.1 is the minimum necessary. Tom Holman is a fan of 10.2. But for most practical purposes we are stuck with 5.1: it is hard enough to get consumers to accept more than two loudspeakers!

I. The Approach

In Your Opinion, What is the Main Contribution of Surround Sound to the Experience of Music?

Morten Lindberg There is no method available today to reproduce the exact perception of attending a live performance. That leaves us with the art of illusion when it comes to recording music. As recording engineers and producers, we need to do exactly the same as any good musician: interpret the music and the composer's intentions and adapt to the media where we perform. Surround sound is a completely new conception of the musical experience. Recorded music is no longer a matter of a fixed one-dimensional setting, but rather a multi-dimensional enveloping situation. Stereo can be described as a flat canvas, while surround sound is a sculpture that you can literally move around and relate to spatially; surrounded by music you can move about in the aural space and choose angles, vantage points and positions. The beauty of the recording arts is that there is no fixed formula and no blueprint. It all comes out of the music. Every project starts out by digging into the score and talking with the composer, if contemporary, and the musicians. It is not our task as producers and engineers to try to re-create a concert situation with all its commercial limitations. On the contrary; we should make the ideal out of the recording medium and create the strongest illusion, the sonic experience that emotionally moves the listener to a better place.

Bob Ludwig Harry Pearson, the man who founded *The Absolute Sound* magazine and helped establish the whole high-end audio marketplace, when defining what he meant by "the absolute sound," said that a live performance was the benchmark to which all reproduced sound should be compared. For me, well recorded surround music, be it 5.1 or 24 speakers can get closer to the "absolute sound." While more speakers, especially the addition of ceiling speakers, yield a more life-like experience, even well done 4-channel recordings can create a sense of reality stereo can never approach.

Jonathan Wyner Presenting music in surround is much more engaging than in stereo. It requires the listener to interact with the sound. A stereo presentation is often closer to mono or a point source in a room where they are listening and doing other things. Surround listening is visceral, engaging and requires the listener to pay attention to a greater degree. Surround gives a wider image, and once one has worked in surround, one realizes just how "cramped" the stereo field can be. Simply by adding an extra 10 degrees on either side of the front image one can realize greater clarity and sonic real estate for elements in an arrangement. It allows for less processing and a more *natural* timbre of elements. Surround gives composers, producers and mixers a wide pallet to enhance the emotional impact of recorded music. Once one works in surround it's disappointing to go *back* to stereo.

Are You Dedicated Primarily to Mixing or Mastering?

Dave Glasser Mostly mastering, but in a few projects we mixed and mastered in the same session. Lately my surround work has been music-centric movies and we've been asked to create surround music mixes and to mix the dialog in surround. So the line is blurred: I like to bring the work in progress to a local art-house cinema with a good sound system to check the mix in a theater environment, in addition to the small-room setting of the studio.

Jonathan Wyner I have been doing some surround mixing as well as mastering. In my mind the line is blurred between the two. Partly because while mastering, I sometimes find myself needing to redistribute the "sound field," e.g., relying on phantom center or creating center channel information, managing the LFE channel or rebalancing energy front to back in order to create a consistent presentation across a project. In stereo, most balance issues are set (hopefully) and there is a clearer separation between the mixing and mastering.

Rich Tozzoli I will never claim to be a mastering engineer. Bob Ludwig did my last DVD project, and he rocked that. Dave Glasser has also mastered a few of my 5.1 SACD discs and he makes a big difference in the final product. We won a Surround Music Award together.

Bob Ludwig Occasionally we have to make changes. I was doing a Coldplay project and they needed a clean version. Unfortunately the instrumental versions still had the F-word reverb bouncing off the rear of the hall. So we had to get all of the tracks to fix it.

If you have to manipulate something, you can stack up groups of six tracks pretty quickly. We worked on a live Foo Fighters project and most of the tracks are from night three, but some were from night two and night one. The DVD sequence was in a different order from what they actually played. By the time you have restored all the applause, you can get 48 tracks or more happening without too much trouble.

I hate when it happens, but there are times we've had to make faux 5.1 recordings. The Police had all their hits remixed for 5.1. There were a couple of stereo demos that ended up being used and the producers wanted to maintain the feeling of 5.1, not have it revert to two channels. I did an Unwrap (TC Electronic) which creates a believable 6 channel presentation and did some further panning. A Brian Ferry 5.1 DVD also used a demo.

Morten Lindberg Our company do a complete workflow from venue recording thru editing, mix and mastering into publishing on the 2L label. We carefully considered the "fresh ears" approach to mastering but eventually decided the consistency from the original recording sessions was more valuable, as the mastering would often benefit from preserving the logic and reasoning in the original recording.

Do You Ever Get Stems?

Dave Glasser The stems that I have gotten were for a live performance DVD, a Grateful Dead movie. The music has one stem and the rest, dialog, behind the scenes elements, which was nice. But usually we do not get stems for music-only projects. For projects that involve other elements beside the surround music, stems: dialog + music are invaluable.

Jonathan Wyner I'm more interested in stems when working in surround than in 2-channel, where I usually discourage it unless it is critically important, e.g., when a client is really out to sea with a mix. With surround, I'm happy to get stems so that I can more easily make adjustments to the width of the soundfield and panning across the front. If I want to spread the image a little wider, it's much easier to control when I have access to individual elements.

Dave Glasser Some people would mix the vocal, for instance as a phantom center but also as a hard center, and so if you do need to do something specific to the vocal, you could always work with the center channel. So depending on how it's mixed, you can work around those limitations. I probably wouldn't want stems un-

> *"Switching from mix to mastering is a mental flip."*

less it was from a producer who really had their act together.

Rich Tozzoli I do six-channel mixes, no individual stems. It ends up as a continuous six-channel mix with all the audience between songs. Then I assemble and cut a sweetened six-channel master of one non-stop performance. However, I also bring some multichannel audience to mastering, where you can fly it in as necessary. Or we often create two six-channel tracks and have one six-channel song feed into another, then "glue" them together in a final layback to Pro Tools or a video deck.

Dave Glasser Actually I might prefer that if somebody was mixing totally in the box, which more people do – why not bring the whole session in? It could be a can of worms, but if it's a producer or an engineer who knows what they are doing and who has a sense of perspective, it might be easier than doing stems, as long as the producer understands this is not another mixing day. But if you envision "a little more vocal," or "I need to pan these guitars back a little bit further," then you have that option.

Morten Lindberg And in fact our sources for mastering are either rendered mixes or the complete open project for more versatility.

Bob Ludwig We seldom get stems in 5.1 except on large live shows where we get sent the entire session as it was mixed "in the box."

Are Most Mastering Sources Pro Tools Sessions?

Rich Tozzoli They're always Pro Tools except sometimes we print right to HD decks for broadcast, as they can take eight channels of audio. I'll bounce to disc or I print right to the video deck. Most often I do not attend the mastering, and if there is no mastering I send the mix files via FTP directly to the production house for encoding and layback, or if there is mastering I FTP the files to mastering. I may send the Pro Tools sessions as a safety. If some little problem should arise, you can clean it up real fast.

Morten Lindberg Our own projects are all done on the Pyramix system from Merging Technologies. External work is nine out of ten times originating from Pro Tools sessions. Complex projects with lots of plug-ins are imported to Pyramix as rendered audio files. For straightforward classical projects we prefer AAF import that makes it possible for us to contribute with a final check of cross-fades, working directly on the originally recorded audio files.

Dave Glasser Projects with a video element inevitably are ProTools sessions. For music only projects, I prefer to work in soundBlade (for PCM), or Sonoma (for DSD).

Bob Ludwig Almost all of it is done from Pro Tools.

What Do You Discover When You Hear Your Mix at Mastering?

Rich Tozzoli You actually hear positive information and the engineer helps your mix, as that is what you go there for. I find a good mastering engineer makes things 20% better. I am very careful with my mixes, I QC them in headphones before they get to him. That is another tortuous process because an hour and half concert takes several hours time.

Morten Lindberg Even though we keep the workflow within the same room, switching from mix to mastering is a mental flip. To me it means zooming out and considering the total emotional impact.

Does a Surround Mix Take Longer to Do Than Stereo?

Morten Lindberg Depends on how it has been produced and recorded. Our approach is that a good surround sound is not created in the mix. It is made in the recording with dedicated microphone techniques. The composers and musicians should perform to the extended multi-dimensional sonic sculpture, allowing more details and broader strokes. Then surround is just a matter of opening up the faders. When the music is created, performed and recorded for surround sound, then stereo is our most time-consuming challenge; figuring out how to preserve the total impact and level of details from this sculpture into a flat canvas.

Rich Tozzoli Yes, surround takes longer because there is less masking – which makes it much harder. Surround brings out imperfections that you didn't know were there. So, if there is a noisy channel, you can't mask it. When I did David Bowie's Ziggy Stardust live with Tony Visconte, the 12-string was clipping on the original 16-track analog master from 1973, which was mostly masked in the stereo mix. In surround, it was very clear on the center channel, so we decrackled it with Waves – carefully judging the quality loss.

Jonathan Wyner It really depends on the type of presentation. In cases where we are simulating a concert hall sort of experience, it doesn't take so much longer to set the soundstage. As Rich points out, however, QC does take longer by a factor of at least two. When mixing in a way that takes advantage of creative possibilities it can take much longer given new possibilities for placing instruments and effects. Dealing with noises… that is a fascinating question. They are potentially more distracting in surround, depending on where they come from. When the noise comes from behind as it might for instance in the audience during a live concert, it has the potential to be a great distraction to the listener. What we try to do is keep people in the illusion that we're creating. The whiplash effect, or whatever people call it, can be extremely distracting [BK: Tom Holman calls it the exit sign effect]. The point is that it is harder to hide problems in surround, and we work hard to minimize flaws and distractions. Some of our newer denoising tools are evolving to address surround.

When mixing, I usually start with the stereo mix and then unpack it into surround. Emotionally it's much more difficult to go back to stereo after hearing the surround mix — so disappointing to lose all the envelopment you get in surround.

BK: I mixed my last surround project first in surround, then when remixing for stereo I used all the tools and techniques at my disposal to try to retain as much as possible the depth and dimension I had achieved in the surround. So the surround acted as an inspiration. Of course the stereo is a disappointment — it can't even come close, but I wouldn't mix the stereo first. For example, the surround version inspired me to try to spread a stereo element wider, or make it more ambient with a dimensional reverb, until I would quickly reach the limit of the 2-channel medium, or the stereo mix became defocused.

Rich Tozzoli In my approach I mix the surround first, followed by the stereo, because that reveals the most imperfections. While working in surround, you set your reverb, EQs, compressions and overall levels. Then it becomes much easier to do the two mix. A separate, independent two mix is not a folddown.

"QC in Surround takes longer by a factor of at least two."

You can then get a stereo record done in a day and a half to two days because everything is already set. It becomes a tweak session. However, the stereo is very much of a letdown compared to the surround mix.

Bob Ludwig I would like to point out that Rich and Morten are the exceptions. Most engineers with whom I work do stereo first and then spread it out for 5.1.

Would You Call "Getting the Perspective from Song to Song" Mastering?

Rich Tozzoli Yes, especially with a surround concert broadcast (Dolby E-delivery) or a DVD concert video, which is a huge market. Audience cuts become so revealing that a whole sweetening session often happens after the mix. So, that is the combination of mastering and sweetening at the same time, which is an absolute art form in itself. When editing, if there is a noise in the surrounds and you have to make a cut, the unmasking reveals a lot of sounds. Often, you have to fly in a stereo audience pair to cover it up. So, what we are trying to do is cheat in and cheat out. Even if you do the finest live recording, you still sweeten the audience levels to make it more exciting. In surround it is that much more complex.

Luckily, I am able to ask a lot of engineers before they go into the live recording, "please put up boundary mikes, audience mikes, balcony mikes." Give me six channels of audience because sometimes they are going to overload, sometimes there are going to be people screaming at one mike. We may have to take out individual claps, as sometimes you can hear one annoying person clearly in surround. And that is something you might not want to spend time on at a costly mastering house. It is definitely what we call sweetening. Unfortunately few people have the money to take it beyond that process for the final mastering step because the budget is in danger. You sometimes have to beg to get mastering money in the world of surround. That is why projects like Blue Oyster Cult were not mastered, which absolutely should have been. They cut the money off and you beg, "Aah, just that one more little bit for the mastering." "Nope! Print it, it's going to tape." Unfortunately, that is what happens .

Jonathan Wyner When going from song to song, the most obvious thing that comes to mind is dealing with perspective vis-a-vis the center channel and sub-content. More often than not, I find that the amount of program that's located in the center channel varies wildly from piece to piece. This is especially true in pop music recordings where days, weeks, or months elapse between mixes. The center channel needs to be consistent enough to present a fairly consistent listening experience across an entire record. A consistent balance between phantom center and center channel needs to be struck. Typically I lean towards relying on a phantom center, and use the center channel to shore it up, to anchor the image.... consider the off-chance that a listener doesn't have their center channel up! It is something that mix engineers need to think about when they are preparing their mixes for surround. If you rely wholly on the center for a vocal, you might open up a can of worms.

We have different strategies for dealing with this sort of adjustment. Sometimes it is simply a matter of adjusting center level to side channels and sometimes it requires a little judicious midrange EQ. It's usually a subtle adjustment in order to get a sense of consistent width across an entire record. There have been instances where there was no phantom center during some surround mixes and so I matrix a center channel.

Dave Glasser With stereo, every now and then you end up re-balancing the channels because maybe it's a little left heavy. With surround, we're re-balancing the channels quite often. Occasionally you cheat things by taking the front and bringing them into the room a little bit more. I agree with Jonathan that the balance between phantom center and hard center is critical. There is a sweet spot where the soundstage opens up – too much hard center information and the soundstage narrows. Talking to mixers, I've concluded that the best 5.1 mixes start with no center channel – get a good mix without the center then add elements to the center channel to anchor the image. Same with the LFE channel.

Bob Ludwig To me, just as any stereo music needs mastering, the 5.1 needs mastering. On rare occasions we'll get something that comes in that's so good that it doesn't need anything. I've mastered some of Tom Jung's – five microphones direct to DSD, where it called for just re-panning some of those five microphones ever so slightly. It was such minutia.

Morten Lindberg Could be a factor in our mastering, but most important is the core sonic values of each track. This is especially important in a consumer market where the album format is broken into single tracks and re-distributed in playlists.

II. Surround Monitor Quality

What Level of Monitor Quality is Required?

Bob Ludwig A mid- or near-field speaker might be fine for a mixing engineer, but you would never want to master with those. We put a lot of effort into the studio and gear. I have Eggleston Works Ivy and Andra speakers, top-of-the-line Transparent Audio speaker cables and interconnects, bridged Cello Performance Mark II amplifiers capable of putting out >4,000 Watts driven by analog preamps that have 120 volt DC rails. They live in an excellent room, it is a joy to hear every day.

Morten Lindberg A pure signal path and high quality full-range monitoring is our most important tool in both mix and mastering.

Jonathan Wyner You want something that's truly full range for mastering. Without it you are essentially flying blind. The quality of mixing speakers follows that which is typical for good mixing studios in stereo. Speaker placement is key. I've encountered problems with mixes done using dipoles in the rear that result in too much presence coming from the surrounds.

Bob Ludwig It's also a function of how well they know the speakers. Look at all the great mixes I got from Bob Clearmountain. They were done on NS-10s. You could also look at the ITU recommendation for stereo monitoring [links at digido.com].

Dave Glasser It's no different than stereo mastering – we want the most accurate and revealing monitor system possible. But I also check the master on a bass-managed system, and sometimes in a small cinema as well. Since surround is new territory for many clients, I encourage them to come in with some mixes or send some surround mixes ahead of time. One problem that I come across is where the front and rear speakers are not integrated very well. I think it is mainly because of how they have their system set up and dialed in.

Rich Tozzoli I mix on NHT Pros at home. The reason I prefer them is they are the perfect blend of what the consumers can hear, which is ultimately what we really need to listen to, and what I need to hear as a music

"*Without high resolution monitoring it is difficult to make decisions no matter the skill of the engineer.*"

professional. I use full range Genelecs and a sub when working in New York City.

Are Mastering Engineers Becoming Dinosaurs with High Resolution Monitors?

Jonathan Wyner I don't think so and might even suggest that the opposite is true. The need for quality control on every level is very high and the acoustics of mixing environments is in a state of decline. Mixers can't always afford to mix in proper studios. Mastering is the last chance to catch problems! There are sometimes elements in a recording that are not necessarily audible in most control rooms or to consumers, but they have implications for what the consumer hears regardless of the resolution of the delivery format. Low frequency 'P'-pops, for example can have a fundamental frequency well below what most speakers might reproduce, but, if there is an AGC circuit somewhere that gets a hold of that and pulls it down, it can create a jarring effect for the listener... and if it's not dealt with, one may find out about it later the hard way, at the most embarrassing moment. If there's a problem you've got to know it's there. You don't always have to 'fix' it—distortion may be acceptable to a producer or artist, but you've got to know it's there and let artists sign off on it. That's part of our job as mastering engineers.

I'll take this a step further by saying that without high resolution monitoring (including the speaker-room interaction) it is difficult to make good decisions no matter the skill of the engineer.

Bob Ludwig In a way I have always wished that recording engineers made *such* good recordings that there would be no need for much mastering after the recording. While a well engineered classical recording often doesn't need anything done to it, with pop productions being recorded and mixed more and more in basements done by friends, there seems to be more of a need for fine mastering now than ever.

Morten Lindberg It is often quoted that "if mastering reveals a serious problem, you should direct the process back to a new mix." I would take this one step further and reflect back to the recording. The concept of tracking sessions to capture and harvest are totally meaningless to me. The recording session is where the music is created. This is where the fundamentals of sonic experience originate. The main values of mix and mastering should be to understand, nurture and enhance that original concept. Coming back to the question; yes, in mastering we really need those high resolution monitors to reveal any flaws in the near-finished product. I believe Bob Ludwig has captured the perfect company name for his services. The music industry more than ever needs observant gatekeepers to preserve the quality of our craft.

Do You Have a Surround-dedicated Room?

Dave Glasser My room was designed by Sam Berkow; we designed it so that it would work for surround. Most of the time, it is used just for stereo, but the surround monitoring system is permanently set up. When people come in to do stereo, they usually ask, "Is the center speaker on?" and they say, "I swear I hear sound coming from that speaker." As for left-right angle, I find that 60° works well in stereo with my Dunlavys, which people have always said work better spread a little wider anyway.

Morten Lindberg Yes, we have a 60m² dedicated 5.1 mix and mastering room.

Bob Ludwig We have Adam Ayan's room and my room, but only my room is set-up for 5.1 mastering. We have a Z-Systems 256 x 256 router that can re-patch the gear with a mouse click from stereo to 5.1.

III. Level Calibration and Loudness Normalization

Let's Start with the Surround Levels. Are you Mastering with the Music Surround Calibration or 3 dB Down (as in Film)?

Dave Glasser Music. Equal level.

Jonathan Wyner The same.

Bob Ludwig Equal level unless it is for theatrical release where the rears are a row of speakers instead of a single point source. It often seems there are film people for movie theatres, home DVD people, and broadcast people who don't realize there are different needs for each.

Rich Tozzoli We are using equal level surround calibration even for concerts with video. In broadcast, that is the fine line again where you will have to mix for the consumer. Bob turned my surrounds up 3 dB and compared to film practice, I usually print them hot. It was definitely the right call.

Morten Lindberg Our monitoring is calibrated 1:1, including the LFE.

BK: This brings up a good topic, since there are two standards for the surrounds and two standards for the LFE. Producers creating home theater masters from film soundtracks must turn down the surround tracks 3 dB. Likewise, producers working in music have to deal with two LFE standards: The Sony/Philips SACD standard aligns the LFE 1:1 while all other media align it at +10 dB. To my mind, producers creating music masters for both SACD and the other media should align their LFE channel 1:1 in their monitors during production. This is functionally a valid SACD master. Then when transferring to the master medium for BluRay, DTS, and so on, they should turn down the LFE track in the master by 10 dB. If you work in the other direction, starting with a PCM-based mix, you risk overloading the LFE channel when turning it up for SACD, though in typical music mixes, the LFE is rarely hot enough to risk clipping even with 10 dB added gain.

Are You Doing Full Level or Normalizing with R-128 or ATSC?

Morten Lindberg We're working full-level peak normalized. But loudness normalization is on our radar and we're preparing our workflow for the EBU R-128. But there are some obvious challenges to this subject. All the way back, our approach to dynamics has always been sort of the loudness method as now described in the R-128 in the sense that we don't "crush" the dynamics. This is somewhat in the tradition of the musical styles we work. Differentiating from some of our more conservative colleagues we do make some manual dynamic envelope control in our mastering process, even for classical music, but not nearly as much as is customary in rock and pop. In practical terms we would hold back 3 dB on a full orchestral dynamic, adapting to a more convenient home playback environment. [BK: But still, Morten's recordings sound "very dynamic" to me when heard in my mastering room, so I think his practice is working out well]. Where we deviate from the R-128 is in the fact that we do apply peak reference level rather than the loudness reference level to the final master files. I would highly welcome a transition to loudness normalization but I'm not convinced this should be introduced on the content delivery level.

When we come to the reference level, this is actually more of a playback issue. It doesn't really matter what is actually the level in the PCM file. LUFS should be analyzed and provided as metadata. Then it is up to the playback system to apply the reference level alignment.

I think the method of measuring LUFS can be applied to music, but already at this point we see that iTunes and streaming services aim for different reference levels than broadcast has decided. There are also some serious flaws in the R-128 algorithm with regard to surround sound. Comparing a conservative front-loaded 5.1 to an immersive 360-mix, the LUFS does not provide a consistent listening experience. The most important aspect of the R-128 is that it removes the commercial motivation for crunching dynamics in mix and mastering.

Dialnorm and Consumer Dynamic Range Compression

BK: **Dialnorm** is Dolby's normalization standard, based on dialogue level.

Bob Ludwig When we first started doing DVD video before DVD-A or SACD was invented, DTS was a big player. DTS really got surround off the ground with their 20-bit surround CDs. When the DVD video format was invented, they pushed to have DTS as part of the specification. In 1997, when we were doing the early surround discs, you had to have something in the LFE or the demo person felt, "I just sold them a $60,000 system, and they can't hear anything out of the LFE. They're going to get worried."

Back in those days when consumers compared the Dolby Digital with the DTS in order for the Dolby to sound as loud as the DTS you'd have to increase the dialnorm to unity gain, -31.

DTS does not have any down mixing nor the "Dolby compression" option that is the bane of my existence. Lots of old Bose systems default to "compression on" unless you purposely turn it off. We did a live DVD and everybody approved the references. The producer of the video went out to the DVD plant to approve the run because he was behind schedule. He called saying "the sound's not right out here." I said, "it's got to be the Dolby Digital compression." He said, "Oh no, it is not that. They told me that it's not on. Plus every player I play in the facility has the same pumping sound." To make a long story short, apparently, up until that day, every Dolby system in their place had the compression turned ON in their decoders! He turned it off and said, "Ah, that's what I remember." So, they actually stickered those particular DVDs, "For best fidelity, in your DVD player's set up menu, set Dolby compression to *off*."

When we first started authoring, all discs defaulted to 5.1 because everybody was trying to push 5.1. So, we used to do dialnorm -31 as a rule. But if we author a disc that's got dialnorm at -31, now we have to make it default to stereo, because if you default to 5.1, you might hear a folddown instead of true stereo unless you take the time to go to a menu and change the audio settings, which a lot of consumers don't bother to do. If somebody listens to a folddown, the Dolby downmix compressors will just go nuts. You never want the listener to hear that. So, we defaulted to stereo so they would hear the dedicated stereo mix, which is usually a PCM stream or a pretty good Dolby 2.0.

This is an authoring, not a mastering problem. You need to have a dialogue between the authoring place and the producer. When James Guthrie did Dark Side of the Moon, he set dialnorm to -24, which is 7 dB lower in level, because that way it is more universally playable if somebody leaves the Dolby compression on. It is a workaround, which is very sad. With so many Dolby chips set to have the Dolby Compression on permanently my current suggestion is to use dialog norm set for near -24 to be safe.

Jonathan Wyner The relationship between level, dynamic range, and delivery to the consumer is such a morass and impossible to generalize about without having the context of a specific format. In general we will run level as usual and then adjust to suit the application. I've noticed the variety of sound that comes out of the DACs in players is astonishing even where so-called *audiophile players* are concerned.

Morten Lindberg We've never related to dialnorm or DRC in our productions — maybe that's fortunate!

IV. Monitoring: Full Range vs. Bass-Managed

Bob Ludwig In the control room, I have the full range EgglestonWorks speakers that go down to 13 Hertz all by themselves, plus I have a pair of M&K subwoofers, just for the "point one"; there is no bass management. It sounds just glorious in there. In my client lounge, right outside the door of my studio, is a highly bass-managed Bose Home Theater "Lifestyle" System. We have the movie EQ turned off and the dynamic compression turned off.

Happy to say that when the Bose system is hooked up the way I like, it translates beautifully from room to room. I usually check everything between the two systems. Occasionally I will hear something in the bass-managed system where it treats the bass a little bit differently than I had imagined it. There is a certain range of acceptability for EQ, I'm thinking "should it be a dB hotter or a dB lower at 60 Hz?" Sometimes what I hear on the bass-managed systems will influence that EQ decision. You definitely need to check all 5.1 on a bass-managed system, especially if there are phase problems with the bass.

Cars can be a great place to hear 5.1. In 2006 millions of cars came out with 5.1 for the first time. Unfortunately it hasn't created a lot of consumer demand, much to my dismay.

Dave Glasser Our Dunlavys are not equal sized. I have model SC-Vs for left and right, four ways with 12 inch woofers and model SC-IVs for the centers and the surrounds. But the 4s for all intents and purposes are full range. I have a Martinsound ManagerMax bass manager that I can insert into the monitor path to check how it works on a bass-managed system. It redirects the low frequency information from all five to the subs. And it does degrade the sound of the speakers a little bit because they are not designed to be rolled off on the bottom. But you do get a good idea of how a bass-managed system is going to behave. We do this because we do not have the luxury of another room.

Rich Tozzoli I use the Waves M360 bass management setup. I do mix with bass management. I pop it in and out. You will see channel 6 in Pro Tools, which is the feed to the sub, disappear when you pop bass management out because I barely print LFE. I used to send a lot more than I do now: for example, on the kick, I used to kick the LFE way up so it sounded great. But it was way too much, it was muddy.

But in New York I also run through Dolby AC-3 hardware encoders into a consumer home theater in another room.

Morten Lindberg I mix and master on a full-range system and make sure that whatever is in the LFE is a unique source and not derived from the main channels. This approach makes it rather safe when the program is played back on a bass-managed system. Some would claim that if you have five full-range speakers you could sum the LFE into your main channels, but I find that a dedicated LFE speaker behaves differently and needs to be treated as such.

Jonathan Wyner I work on a full range discrete 5.1 system. Having a satellite/sub bass-managed system for reference is very important to get a reality check.

BK: I would like to supplement the remarks here. A well-known surround mixing engineer has stated that "the LFE channel is absolutely essential for production of live music because 'full range' speakers simply don't reproduce the bottom end correctly. They don't even reproduce the bottom 'B' on a five-string electric bass correctly." I agree that many so-called 'full range' loudspeakers don't go down deep enough, but I disagree with his conclusion that the LFE channel is necessary to carry this off. This is because if a consumer's surround system is equipped with a subwoofer, and that subwoofer is capable of playing extreme low frequencies, then it is also capable of extending the low frequency capability of the main channels. This facility is called **bass management**, which every home surround system has built in. Bass management is simply a crossover for the main audio information to be fed to the subwoofer, in order to gain more headroom or extension when the main speakers are not capable of going low enough. Technically speaking, with nearly all consumer systems being bass-managed, it is not necessary to put information in the LFE *unless* it would keep the main channels from overloading. Rarely is that necessary – for most music programs, the bass fits quite well into the main channels. Regardless, I advise checking any decisions on a bass-managed system to confirm they are working.

Does this mean that feeding the LFE unnecessarily is harmful? I've heard several music surround mixes which in my opinion have put too much bass information into the LFE channel. It might be a problem if the *same* information is fed to both mains and LFE, because the two systems might not be in good phase and the two signals could combine poorly. That's why it's probably better to feed only *unique* information to the LFE, such as extra low frequency information not being fed to the mains. However, using the LFE in mixing as an artificial *extension* of the main system is making the assumption that the consumer's system is not bass-managed. This opens a can of worms, with the probable result being extra unintended bass on the consumer's system. Avoid this temptation, but in any case, check on a bass-managed system. If the consumer's system is properly set up, then there should be no sonic difference between feeding the bass instrument wholely to the main channels versus partly or entirely to the LFE. If the consumer's system has the woofer turned up too far (which is often the case), then I suspect the sound will be bass heavy regardless. My recommendation is to avoid complications. For example, you may choose to feed the LFE:

- as a supplementary "boost" channel of extra low frequencies for esthetic reasons. Check that the extra bass would be satisfactory without the LFE just in case the channel is not present in the consumer's system.
- to protect the main channel levels if they are running out of headroom and he wants to hear additional deep bass. This is similar to the motion picture philosophy of using the LFE as an effects channel.

The LFE is the most misused and least understood channel in the surround business. Political decisions and technical misunderstandings probably explain why there is unnecessary information in the LFE channel of so many music releases. However, there is a bright side: headroom. If you're making a loud mix, you can avoid peak limiting the mains by spreading the low frequency information to the spare channel, allowing you to pump up the total loudness and get a little quality back. That's bass-ackward but I suppose the better choice!

V. Are You Using ITU Monitor Layout?

Dave Glasser I do not use 110°. I am closer to 130°. That is closer to the NARAS recommendations. ITU is 110° but many people think that it does not work that well for music. I think 110° is definitely not far back enough. What we recommend is something greater like 130° or 135°.

Bob Ludwig In my mastering room, where I do mostly rock and pop surround, I have the rears at 135°. At home, where my listening is more classical oriented, I listen at 110° as I feel there is less of a "disconnect" between the front and the back, plus I know that all the European engineers are using ITU recommendations (and note, it is a recommendation, not a standard). Also at work, in one of my production rooms where I have a ProAc 5.1 system set-up, I also use ITU. For our loudspeaker QC pass, having the rears at 110° delivers more acuity in the rears than the more severe angles.

Dave Glasser One of the big problems is that it sounds like there are two different things going on. For instance, take a live concert. With the band in front and a couple of audience noises coming from behind, it is not very well integrated. So that is where tweaks end up often.

BK: With all due respect, I think 135° is contributing to that dichotomy. That's why they didn't place surrounds behind us in the motion picture theater, because to me it is most disconcerting. But I'm oriented more to envelopment than to localization.

Dave Glasser I just know when I had these speakers at 110°, it felt like there was a big hole in the back. But once you work at it, you can get it to work really well. So I think it is due to people not thinking in terms of creating a sound field. They just see five speakers and decide – okay I'll put something there and I'll put something there, but they're not perceiving those five speakers as one integrated sound field. And that holds true whether you are doing an ambient rear production or whether you are going all out with discrete sources.

Morten Lindberg Our mastering studio is set up and calibrated to the ITU-R BS 775 music configuration.

VI. Has the Volume War Invaded Surround?

Bob Ludwig Let me state for the record that there is no need to have over-compressed recordings. Simply turn the playback volume clockwise if a recording sounds too soft.

The flat 5.1 mixes we got from Nine Inch Nails were mixed hotter than I would have dreamt of mastering it. The meters just pegged and never came off the peg. They are trying to get it as hot as they can get it, and I guess it works fine for their music. The mastered version came out at the same level as the mix. With DVD home theater, unlike movie theatres, there is no adherence to the 85 dB calibration standard. Fortunately, most of the people who are doing DVD video work do not have this louder-is-better mentality that some record A&R people have. I have to say that no one is going to tell Nine Inch Nails to bring their record down.

Dave Glasser I haven't had problems with over-compressed 5.1 material yet. I'm not doing a lot of rock & roll 5.1 though. I'm doing jazz, classical, and acoustic music. So the volume wars haven't caught up to that. In the 6 years since the last edition of this book, I think the loudness race is slowing down.

Jonathan Wyner Fear motivates people to compete in a volume war. You've got to stick to your sense of ethics and do what is right and what is best to create work that sounds as good as it can, and if that involves making something that's 6 dB lower than some "Crusher's" idea of what's supposed to happen then so be it. The real question is, how can we keep our livelihoods? Certainly not by producing unlistenable music that's distorted and causes ear fatigue. We need to stick to our guns and produce work that is worth listening to!

One small silver lining in the demise of record labels is the disappearance of a centralized filter, focusing people's attention on any particular body of musical work. There's not so much direct comparison these days and audiences are getting more different kinds of music from different places and having more different kinds of listening experiences. In some cases that allows us to relax levels. I do not think artists are generally dealing with program directors or A&R offices set up with surround systems who are judging one thing better than another because it is louder. Surround just hasn't penetrated that part of the market place.

I have some clients for whom it was important that the surround version of their project have a little bit more impact and be slightly louder than the stereo version, and this is just in terms of SPL in the room. Fortunately, having the multiple channels in the room provides enough additional SPL that you can usually relax the average level of each channel and still make enough of an impression compared to the stereo presentation. It doesn't take a lot for a 5.0 version to compete with its stereo companion, simply because there is more energy from the 5 speakers.

My contention is that something that is well recorded and well mixed, sounds loud naturally. I personally try not to adopt a practice simply to compete, but focus on the values, ethics and goals manifest in an artist's work.

Morten Lindberg The volume war has not invaded surround partly due to the fact that surround sound is not into the radio race, but mostly because the nature of the medium provides so much more torque and raw power in the 5.1 playback situation. You could paint an analogy to cars; stereo is the Fiat with the 1.2 liter engine where you constantly push the engine speed and 5.1 is the BMW 3.6 liter providing a larger potential.

VII. Sample Rates

Do You See Any 96 kHz Surround Material?

Bob Ludwig Oh! Totally, we get that and 88.2 kHz and even 192 kHz. The Nine Inch Nails was at 192 kHz. In the past we seldom got 48 kHz masters. Now, where things are pretty much DVD-video oriented, I have noticed more projects at 48 kHz. When we were doing SACDs and DVD-As, it was all 96 kHz. These days it is almost always either 48 kHz or 96 kHz although I recently did a Beyoncé surround sound project with videos designed for iTunes and that project was mixed to 44.1kHz.

If a project comes in at 96K, we'll master at 96K and then we will give them six broadcast wave files at 48K for

Dolby encoding. We've still got the high resolution 96K files for a future DVD-A or Blu-Ray. If they make files for digital downloads, they can encode down from that.

Rich Tozzoli I do. Mostly for audio-only projects. 96 kHz usually ends up on a SACD. [BK: Or the Pure Audio Blu-Ray format]

Jonathan Wyner Program comes in at all sample rates. 96 and 88.2 are fairly common and then we'll derive what we need for delivery from that.

Morten Lindberg Yes, most external works now come in at 96kHz. For our own projects we make the recordings in DXD (352.8kHz/24bit) and preserve this resolution all thru editing, mix and mastering. We now also distribute DXD to end-customers in the specialized Hi-Fi consumer market.

VIII. Destination Media

Is SACD a Dead Medium? What Do You Foresee for this and other Future Surround Formats?

Jonathan Wyner SACD as a format has been relegated to a niche market but it is a single manifestation of DSD. However, sigma delta is in play everywhere and I fully expect to see new physical and online consumer formats based on it in the coming years.

I am especially intrigued by developments in playback systems. There are markets where all audio is broadcast in multi-channel although most consumers still hear the stereo presentation but there is clearly an appetite in the marketplace for experimentation. We have seen uptake in the car playback systems and the proliferation of the 'soundbar'. I'm no futurist but it's not hard to imagine that systems that project sound into space without the use of speakers and cables would come into the consumer products. Should that occur I think we might well see an explosion of multichannel immersive playback that we have fantasized about for decades.

BK: Inexpensive combo players from Sony, and premium combo players from Oppo and others are capable of playing SACD as well as any other medium.

Morten Lindberg I wouldn't say dead, but in reality a limited market in decline.

Bob Ludwig I don't think SACD is dead, it beat DVD-Audio in the format war. In Europe, there are still a lot of indie labels that do hybrid SACDs. I listen to the BBC-3 "CD review" program. Often the new releases they highlight are from labels like Telarc, Delos, Hyperion, Pentatone and Harmonia Mundi and they're often hybrid SACDs. That makes so much sense to me. As a classical buyer, I will buy an SACD before I buy a stereo CD. When the first edition of this book was published Arkivmusic.com listed 1,461 available SACD titles. In 2014 I see 3,026 titles. I'm telling you, it's not dead. I mastered the new *Band of Horses* live recording and it is available online as a DSD download!

What About Blu-Ray and HD?

Dave Glasser In recent times, pretty much all of our DVD projects have also been released on BluRay, with uncompressed 24 bit audio. Pure Audio Blu-Ray is also a great format.

Jonathan Wyner I think these could be viable formats but the overall success will not be driven by fidelity… Let's think back for a minute to the history of consumer formats. At what point in the last 40 years has any format had success in the wider marketplace based purely on the improved fidelity where it didn't *also* represent an obvious advantage in terms of convenience or economy for the consumer? In terms of fidelity gains, as the internet infrastructure improves so that the majority of consumers have much faster and wider connections, I am hopeful that 44/16 resolution, if not higher, will become the *minimum* de facto standard. Streaming high-resolution surround seems very appealing with disc based formats having their place. Blu-Ray is a great container for hi res and multichannel audio. Authoring was a bit of a thorn in the side of music producers until recently. If someone sees a market, it's possible to make great hi-res surround products without compromise.

Morten Lindberg With the concept of Pure Audio Blu-ray we now reach a large home entertainment audience where music, gaming and video is melding into one healthy consumer market. And we offer surround downloads in FLAC format as well to this audience.

Bob Ludwig A lot of the old SACD and DVD-A titles I have mastered have been put out by Universal in Blu-Ray. Basically we master surround in as high a resolution as makes sense for the project and the record companies do whatever they wish with it.

What About Just Plain DVD for a Surround Music Release?

Morten Lindberg With the successful penetration of Blu-ray, DVD is no longer attractive to us. The only exceptions are the projects we work with the automobile industry.

Jonathan Wyner If you mean DVD video, it's a good format from the standpoint of compatibility and for presentation of music concerts with video. But the audio spec leaves much to be desired since we're either dealing with stereo PCM and AC3 surround *or* living with folddown stereo.

Rich Tozzoli I always default to the fact that most of my mixing is for broadcast. I feel very strongly about the world of HD television for the delivery of surround. Anybody who thinks they are going to do surround mixing for audio only is sadly mistaken, as it will not happen.

Jonathan Wyner I think the primary context for working in surround will be audio married to video and in the game market, although there will continue to be niche markets representing all kinds of interests: high res, audio only, etc.

I can't get too worked up over exactly what the marketplace will support or is going to allow. It's something I have no control over, so why worry about it? I never saw surround as having a profound impact on the wider marketplace or being something that would drive consumers to a particular format. Example - *Diana Krall: Live in Paris*. Fantastic performance, beautiful video direction, and sound! Al Schmidt did an incredible job. People who are inclined to buy that disc are simply going to appreciate the fact that it looks and sounds great, not that it is surround.

I think it is important for people to understand that the post-production steps in surround, especially when video and authoring are concerned, are more complicated and time consuming then in making a two channel album.

IX. Authoring

Bob Ludwig We were the first mastering studio to offer authoring in December of 1997. We have D5s and digital NTSC and PAL Betacams. We do just as much work in PAL as we do in NTSC. It is relatively easy for us to supply surround clients with any sort of reference.

If you don't have authoring available, you should be able to afford Dolby or DTS software for references. But most of our clients come in with video and want to see references with the video. More and more we simply send back mastered WAV files and the client is checking them that way.

Dave Glasser For SACD, we can provide the finished cutting master. We have contractors who we use for DVD and BluRay authoring.

Jonathan Wyner We provide simple authoring services in-house and can build a DVD video or DVD audio ref. And we have access to world class authoring on the other side of town that can make discs jump through hoops. It is fantastic not only given the proximity and their skill level, but everybody over there comes from an audio world with the exception of one guy who is their video and film geek.

Morten Lindberg For SACD we make the UCMF in-house. When we first considered Blu-ray we looked into the process and found it to be too video-specific. Rather than developing these services within our audio-house, we chose to connect with an external video authoring facility. Preparing file formats and authoring is an important part of mastering but let's not forget that these are mere containers. Our core responsibility is to act as a gateway and quality control of the master source files, for current distribution and for future safe-keeping beyond the short-sighted current quarter revenue reports of today's label management.

X. Reverb in Surround

Dave Glasser We end up using reverb a lot more often than ever comes up in stereo, and I have the TC 6000, which is fantastic, and a Sony S777.

Rich Tozzoli I have also been using impulse responses since they were first introduced with the Sony S777. Once I heard the unit, I recognized the value. As a surround engineer, part of what you can recreate in the world of surround is the acoustic space. Now, I know how to do my own impulse responses and I will use those in the mix because they work. I was very lucky that I got Waves to capture some of my favorite recording spaces, Trinity Church, New York, and I got AudioEase Altiverb to capture Clubhouse Studios in Rhinebeck, NY. So, I have my two favorite acoustic spaces as presets in people's impulse response folders. Software impulse responses make a surround experience much better because surround is about envelopment and realism to me.

Dave Glasser What you can do and what I have done a couple of times is, using the TC – in the front channels, add some early reflections and in the rears bring in the reverb.

Morten Lindberg Surround Sound in its nature provides a rich spatial experience. We use the Lexicon 960 for discrete enhancement. Lately we have explored the PhoenixVerb from Exponential Audio with good results. As a true 5.1 VST supporting high sample rates it makes for an efficient tool within our Pyramix workstation. Consistently, stereo requires a larger amount of added reverb to reach the same level of openness as Surround Sound.

Jonathan Wyner We have a Sony 777, TC-6000 and a pair of Lexicon 300s that we'll use. A little reverb is often the "glue" that adds that last bit of convincing realism to the soundstage.

Benchmark DAC-1. Very solid sound quality. Excellent price/performance ratio.

Crane Song HEDD-192. Excellent ADC and DAC with insert switching for analog devices. Also can simulate the distortion of tubes and tape.

Crane Song STC-8 discrete Class A compressor/limiter. Warm yet transparent sound, very useful for subtle mastering work.
Alternative: The Maselec MLA-4 3-band compressor/expander (page 108).

Crane Song Trakker. While intended for tracking, can serve quite well as a more aggressively-capable mastering compressor. Can also do subtle!
Don't forget parallel insertion via the transfer console if more subtle results are desired. I suggested a longer attack mod which Dave Hill has implemented on later units.

Dangerous BAX EQ. So transparent — when it's set flat you can't tell it's in the circuit. Baxandall-shaped curves with gentle high and low pass filters.
For the final subtle polish but don't forget its nice, punchy bass boost.

CHAPTER 12

Hardware Tools of the Trade

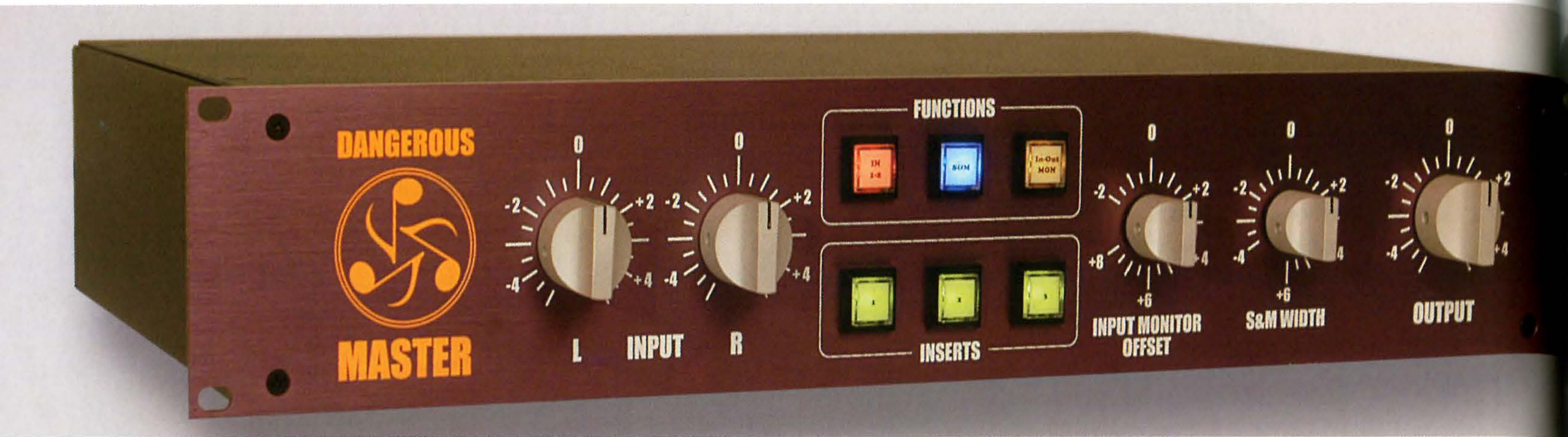
This Chapter is a brief pictorial survey of outboard gear which is useful for mastering. You might ask, "why do you have so many compressors?" I currently have three outboard analog dynamics processors and three outboard digital dynamics processors in the mastering room. Isn't this overkill? There is no "one size fits all" dynamics processor. If you want to be versatile, you probably need *at least three* different compressors, one which is transparent and capable of being subtle, one hard and aggressive, and one which helps to get a lush, "creamy" sound. Remember, "no compressor at all" can be the right solution.

There are far more high-quality pieces of gear than I can include in this survey. Omission does not mean that processor is not suitable for mastering. On the contrary, it may have been omitted due to oversight, lack of space, or simply because I am not familiar enough with it to give it the high sign. Gear lust begins right now!

Anamod ATS-1. Authentic-sounding analog tape simulator with very pure tone. It can be quite subtle or strong. See also the discussion on pages 302-303.

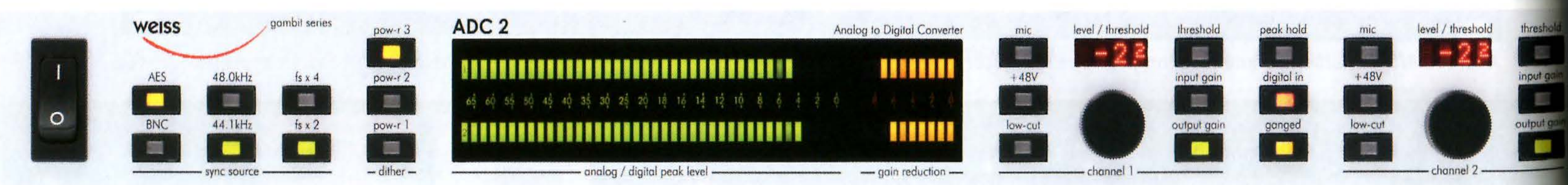
API 2500 Compressor. Great for punching rock music. Can also be used subtly with a touch of an "edge," or even more subtly placed in parallel using a suitable mastering transfer console. "Katz mod" shown, with a maximum makeup gain of 3 dB for a finer resolution of the control. Log settings with a test tone as described on page 344.

Bricasti M7. One of the most natural-sounding digital reverb generators, with excellent early reflection simulation and depth.



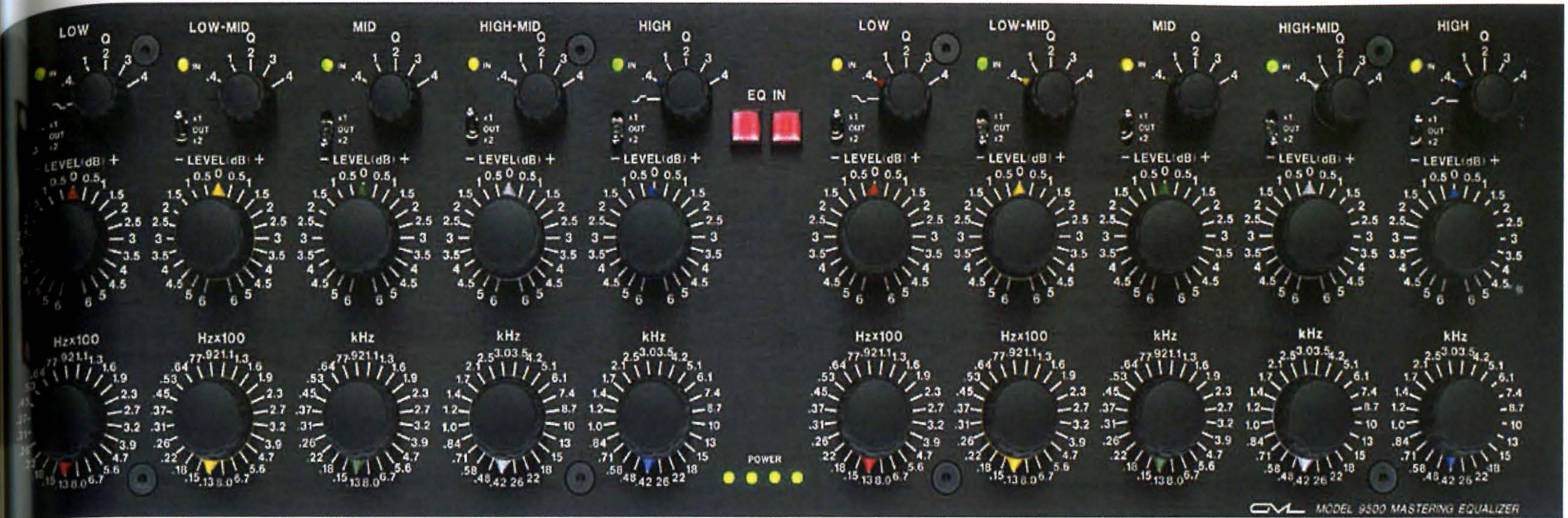
Dangerous Master transfer console. Insert and level functionality with MS width control and valuable source level offset for level-matching mix versus master. Alternatives: the Maselec MTC-1X (page 193), the Crookwood (wide range of versatile units), and the Manley Backbone.

Forssell Technologies MADA-2 ADC and DAC. Very clean, transparent. Useful for inserting analog devices in a digital chain or analog transfers. Alternate: Mytek, Lavry Engineering



Weiss ADC 2. Premium-quality analog to digital conversion. Can also be used as an external compressor/limiter/pow-r dithering unit with digital I/O.

Prism Lyra-2 USB interface, ADC, DAC. Stereo Mic/Line/Instrument input, S/PDIF I/O, 4 channels of line plus headphone output. Premium sound quality and performance in a small, economical package.It uses the Prism CleverClox Hybrid PLL which avoids using ASRC chips that would change the data.



GML 9500 parametric EQ with mastering detents. Very pure-sounding analog equalizer with optimum Q.

Manley Massive Passive Tube EQ. Not too tubey, nice character with unique curve capability.

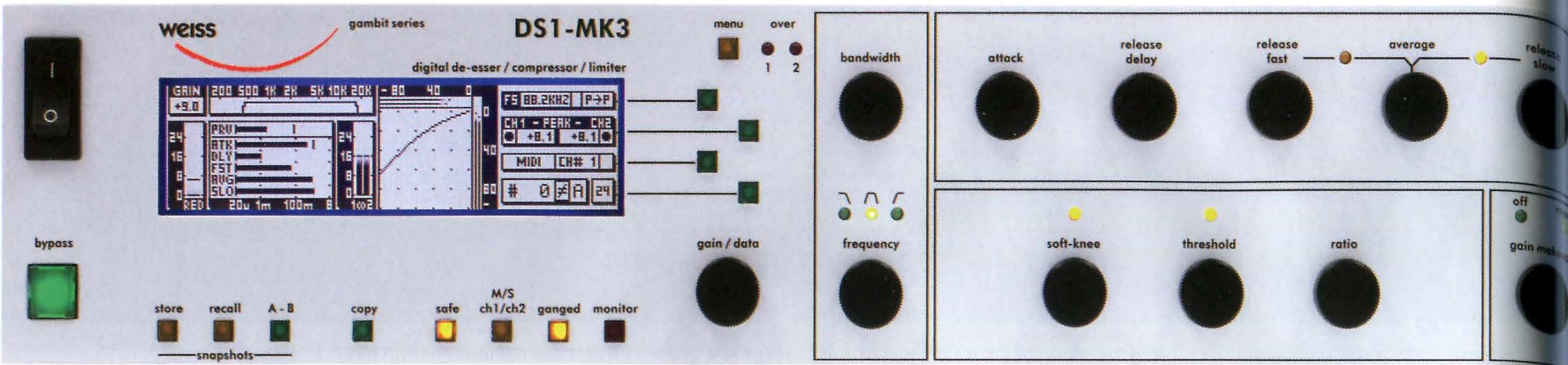
Manley Stereo Variable Mu® tube compressor. Think very tasty sweet cream.

Millennia Media NSEQ-2. Twin topology means we get two sounds in one! In solid state it's very transparent.
Also consider the Bettermaker digitally-controlled analog EQ with memories, described on *page 67**.*

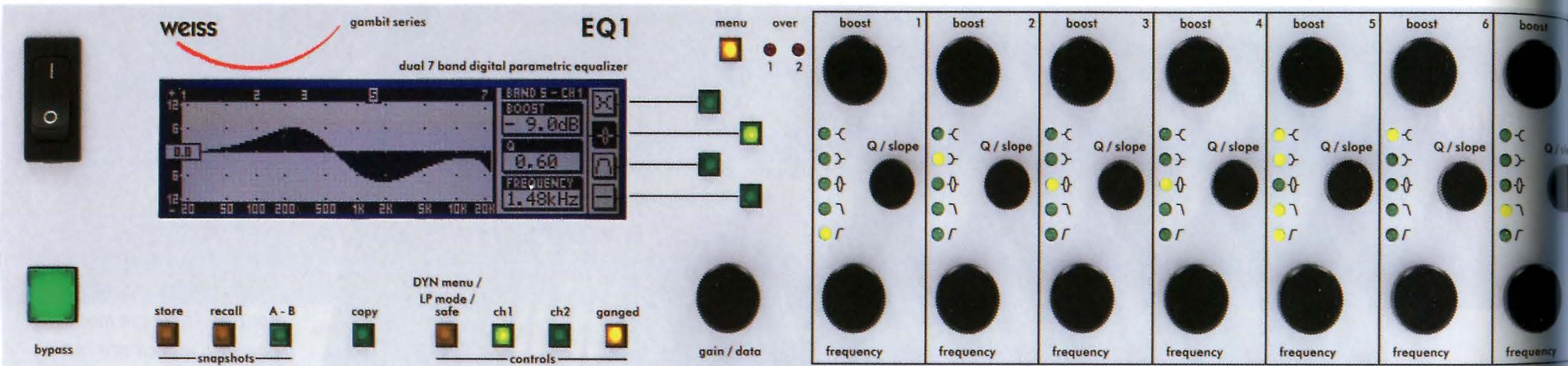
Pendulum Audio ES-8 variable mu (remote cutoff triode) tube compressor with transformer coupled input and transformerless solid-state balanced line output. It's far more transparent than the Fairchild but can be warm, punchy or creamy when you need it, very versatile. See discussion in Chapter 6.

Pendulum Audio OCL-2. Very transparent for a tube unit. Transformerless electro-optical tube compressor which is fast enough for many more rhythmic applications than typical optical compressors.

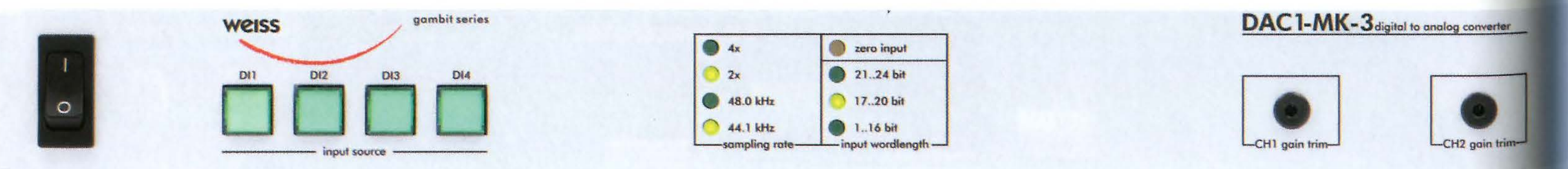
TC Electronic System 6000. This is the Icon remote control. There are more powerful algorithms in here than I can enumerate. I use it for parallel compression, peak limiting, superb reverberation, in stereo or surround. And much more!



Weiss DS1-MK3. Probably the world's best de-esser, especially helpful for mastering because of its invisible action, even on mixed material. When not de-essing, it makes a very nice, transparent single band compressor or expander, surgical when needed. It's indispensable so I have two in my rack!



Weiss EQ1. There are three versions: the standard minimum phase, the combination linear and minimum phase and the dynamic equalizer. Take your pick, they're all transparent, curves from very broad to very surgical. The Weiss units, like the TC 6000, are ergonomic, as easy to use as any piece of analog gear.

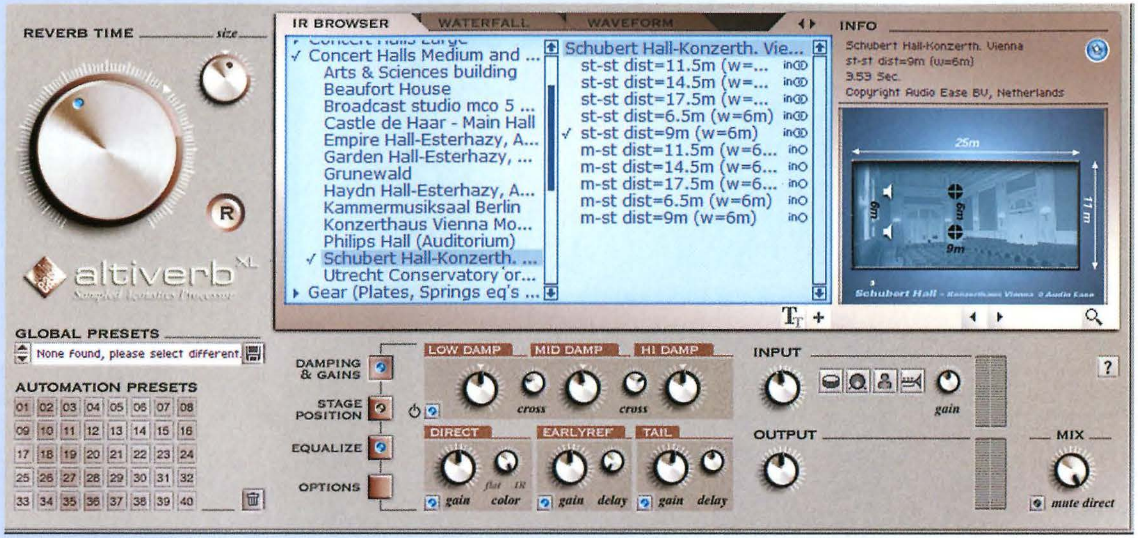


Weiss DAC1-MK3. Premium quality. Four digital inputs. Input selection and level are remote controllable. Extremely low-jitter circuitry accomplished with a superior analog PLL, not an ASRC, so the data is not changed. Excellent SNR.

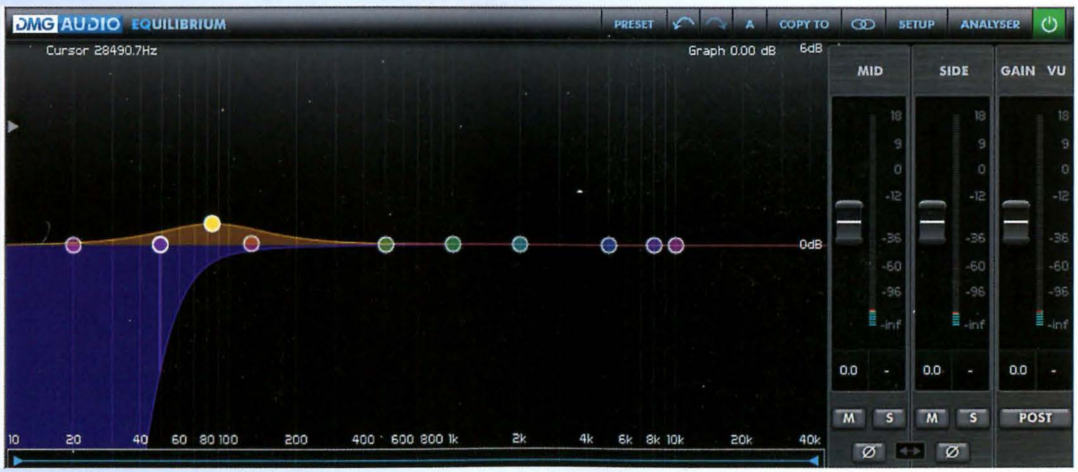
CHAPTER 13

Software Tools of the Trade

Here is a brief pictorial survey of plug-ins and applications useful for mastering. As with hardware, there are far more high-quality plug-ins than I can include here. Don't forget loudness meters (Chapter 18), Bitter (page 210) and restoration tools (Chapter 8).



Altiverb by Audioease. Typically, mastering does not require reverb; it would be like trying to remix and distort the producer's intent. But one of our most common tasks is to fix up tails (page 43). Or, when the client supplies stems so we can use a better reverb than he has available. About 20% of my mastering is from stems as opposed to full mixes. This crosses the delicate border between mixing and mastering, but when needed, one of the best tools is Altiverb.



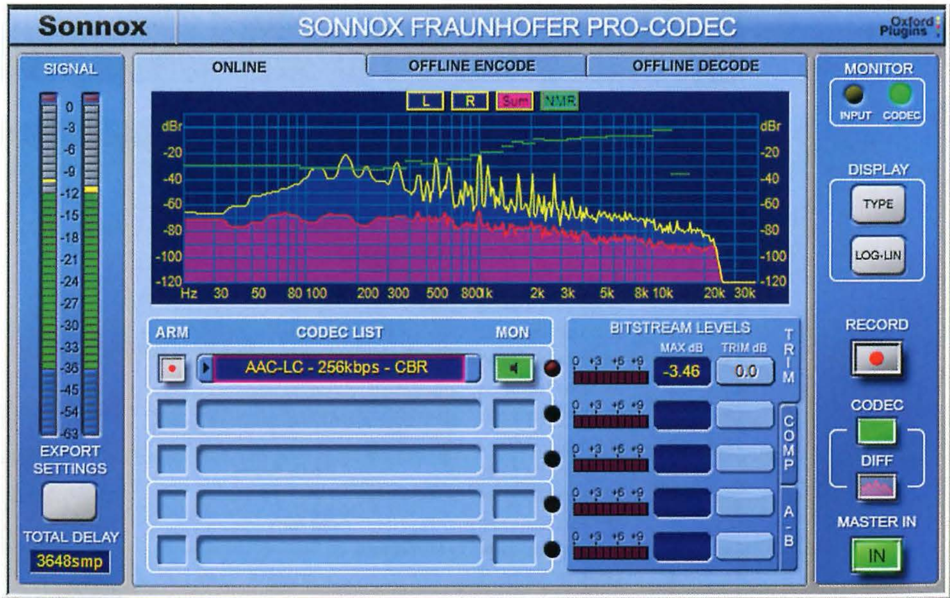
Equilibrium by DMG. Many available curves. Very transparent, high resolution. Any band may be linear or minimum phase.

Ozone by Izotope. One of the few multiprocess plug-ins written with integrity and high resolution. Don't be tempted to use more features than a given project requires!

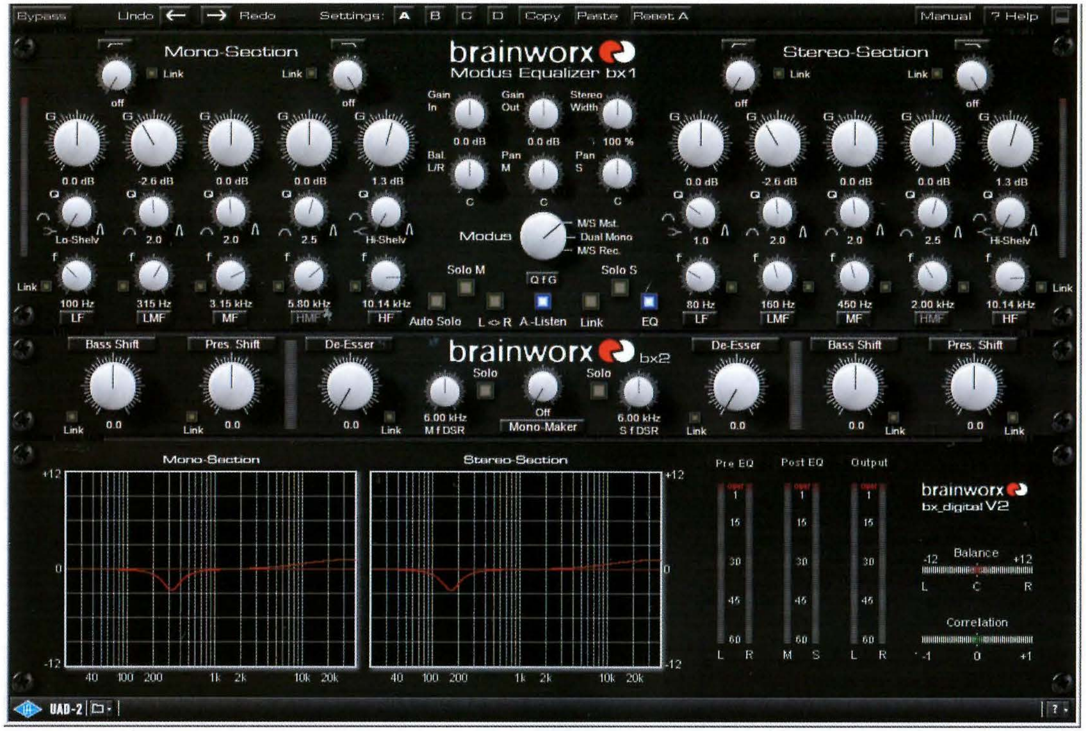
Lowender by reFuse. A real nice discovery. Adds authentic-sounding subharmonics, pure or with distortion as desired. I've used it to supply low end punch on mixed material when a bass drum was missing bottom, fatten the bottom end on an electric bass, supply authenticity or effect when the mix was looking for it. Subtle goes a long way, and be absolutely certain your monitoring is accurate before using this processor — it will supply low end that goes down to the center of the earth!

Speakerphone by Audioease. Mostly I've used this on DI bass stems to help the note definition, supply a sense of space and to sound more "real." It's designed more for mixing than mastering, but I use it often enough and it can sound indistinguishable from a miked loudspeaker.

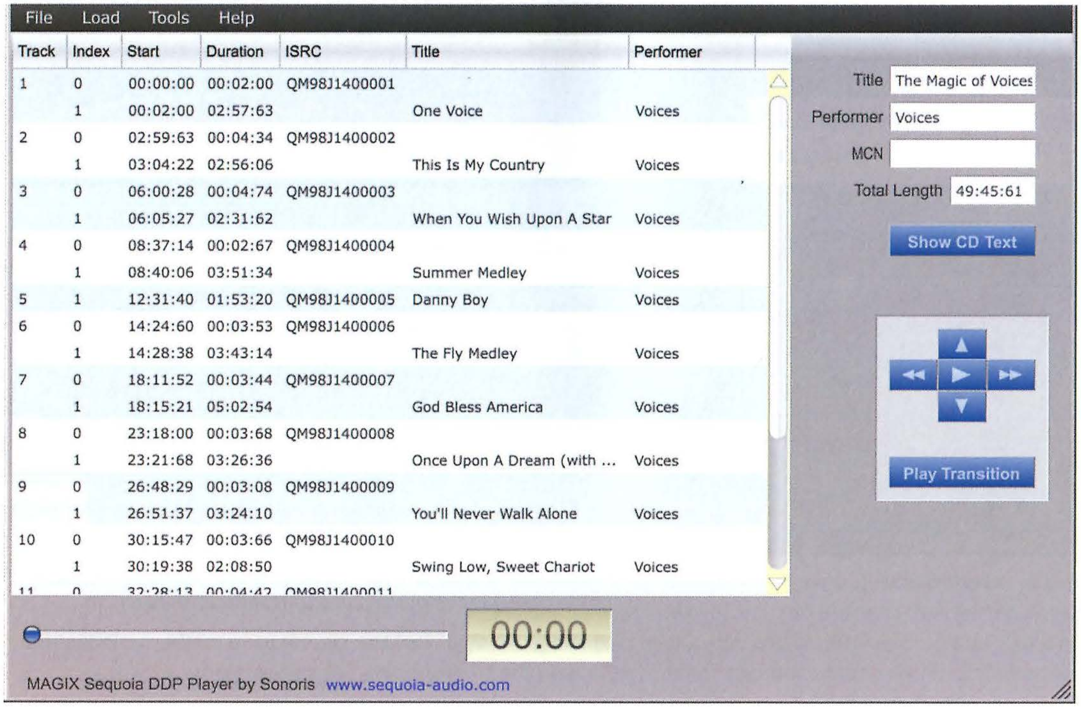
oldTimer by PSP. I'm seduced by the look, but it sounds as sweet as it looks. Mostly I use this on stems.



Fraunhofer Pro-Codec by Sonnox. Demonstrate the effects of coding with any chosen codec or bitrate. On Windows it is not the exact AAC codec used by OSX, but it is close. Also can convert any file(s) between coded and PCM for later comparison in a DAW without the plug-in.



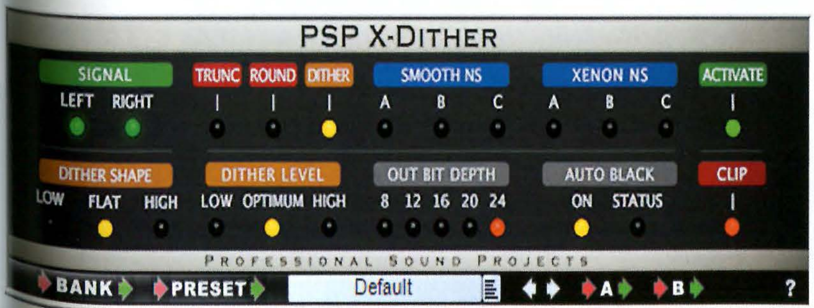
Brainworx Modus Equalizer for UAD. An excellent MS EQ. Alternative: DMG Equilibrium in MS mode.



DDP Player by Sonoris. An essential tool. Provides a secure means for clients to proof and cut their own CD reference. Sonoris can "rebrand" this for any mastering house.

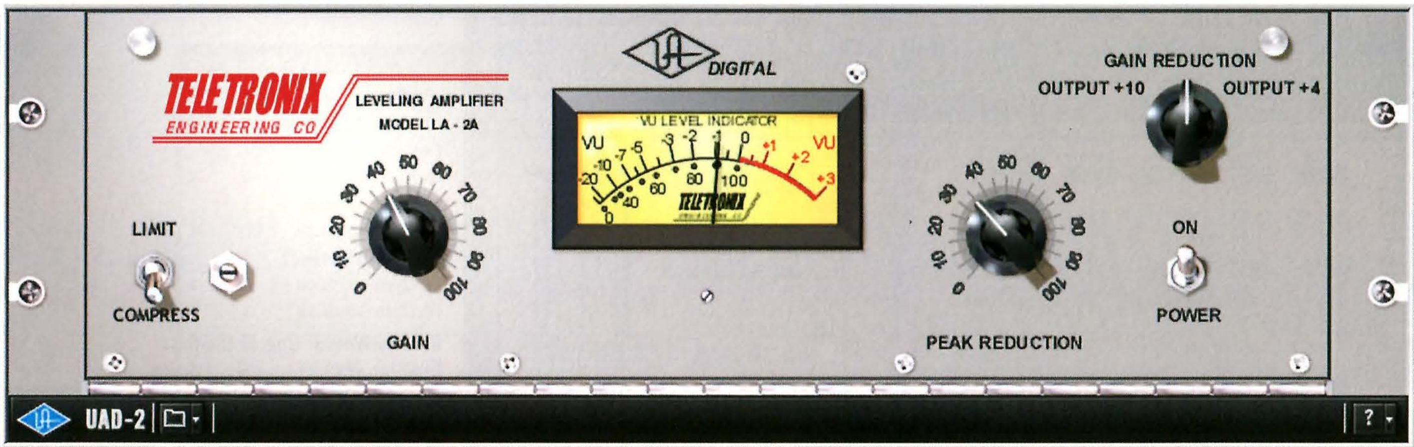
EMT 250 by UAD. In 1976, Barry Blesser designed one of the finest algorithmic reverbs ever made. This is an authentic recreation. Nice depth, purity of tone. Great for producing a natural perspective. Mono in/Stereo out.

IR-1 by Waves. One of the first high-quality convolution reverbs.



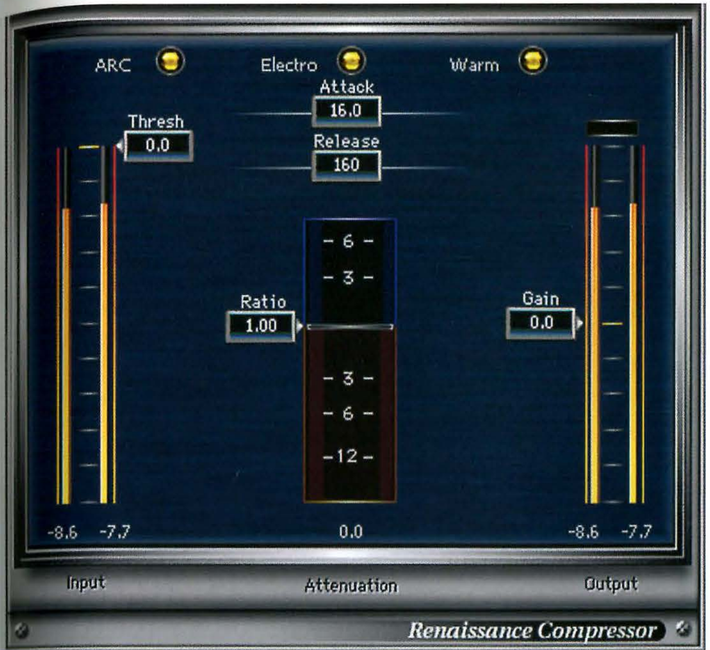
X-Dither by PSP. Shaping tuned to maximize depth and purity of tone in 16-bit mode. This and POW-R are my "go to" 1644 dithers.

Pultec Pro by UAD. Has the classic warmth and those sweet, broad curve shapes.



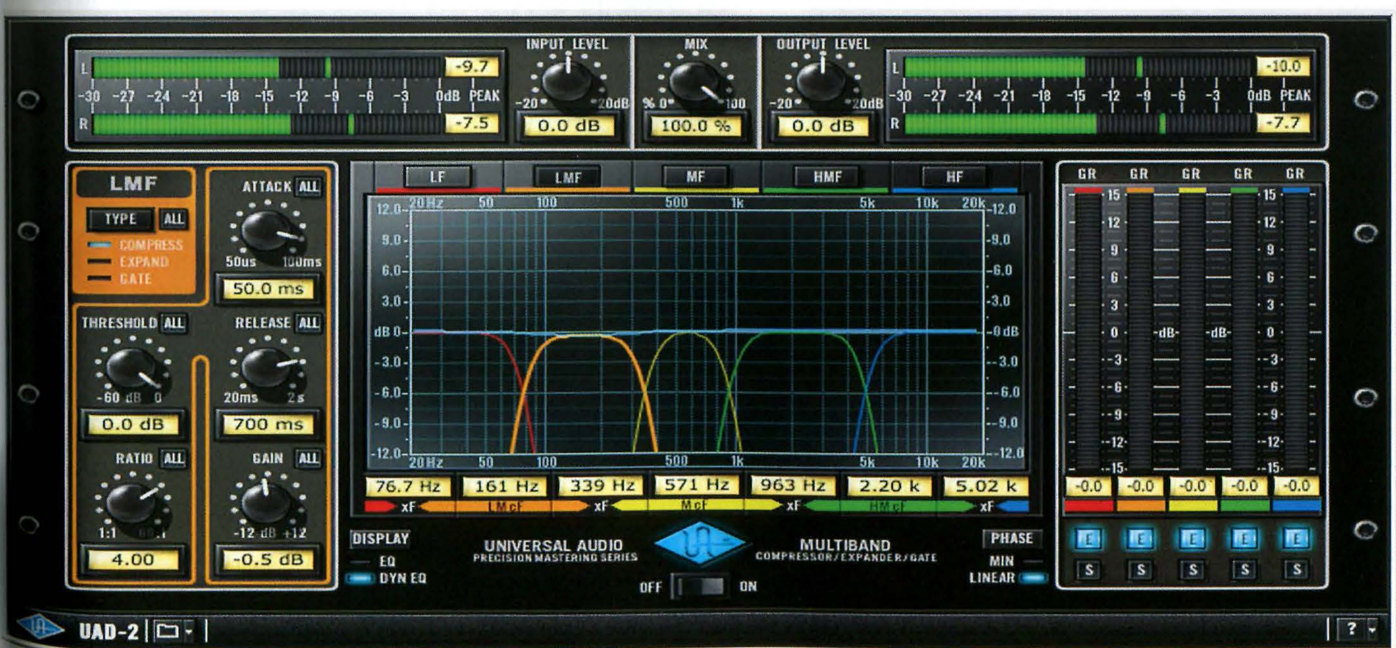
LA-2A by UAD. The original hardware piece is one of my favorites. I rarely use the plug-in on mixes but it's perfect for vocal stems. As with the hardware unit, don't trust the GR meter, use your ears!

Precision Enhancer by UAD. A subtle distortion generator that works quite well when EQ or dynamics processing won't bring out inner details in a problem mix during mastering.



Renaissance Compressor by Waves. A simple processor very suitable for stem mastering when an overall dynamics processor is not enough.

Xenon Limiter by PSP. Versatile oversampled peak limiter with few artifacts if not pushed. This and the System 6000 Brickwall are my "go to" limiters.

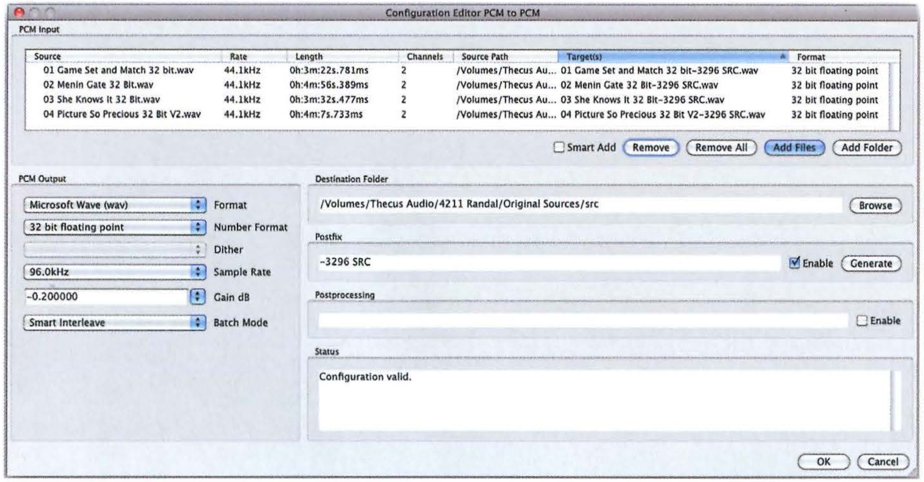


Precision multiband by UAD. Very clean downward compression or upward expansion. I use it to supplement other tools, especially in stem mastering when an element in a stem needs control that cannot be solved by EQ. Don't trust the GR meters, use your ears!

K-Stereo by UAD. Full disclosure: I am the inventor of this processor. Enhances existing depth and ambience in a recording. Subtle is good. Please read the manual.

Ocean Way Studios by UAD. Combination algorithmic and convolution processor with stored characteristics of two nice studios, mike choice and adjustable mike placements. Occasionally useful on full mixes but mostly great for stems.

Ampex ATR 102 by UAD. One of two convincing plug-in analog tape simulators. Leave the wow and flutter off. The Ampex specializes in presence with warmth.



Saracon by Weiss. I'll go out on a limb and claim that this is the world's best-sounding sample rate converter. They say you can pick only one: faster, better, or cheaper. But in this case you can pick two. Excellent batch mode. Alternate: Izotope 64-bit SRC.

Studer A800 by UAD. Sometimes a tape emulator is the perfect solution. I may draw on the ATS-1 (page 170) about once or twice a month. For stem projects when individual tracks need separate treatment I may use the UAD emulations. The Studer has a warm, smooth sound. I agree with Bruce Hensal that "15 IPS/456 is the sound of rock and roll." See page 302 for a discussion on analog tape simulators.

L2 Ultramaximizer by Waves. The original plug-in peak limiter that started it all. Michael Gerzon's algorithms including IDR dither/noise shaping.

PART III: ADVANCED THEORY & PRACTICE

"Making good sound is like preparing good food. If you overcook, it loses its taste."

—Bob Katz

CHAPTER 14

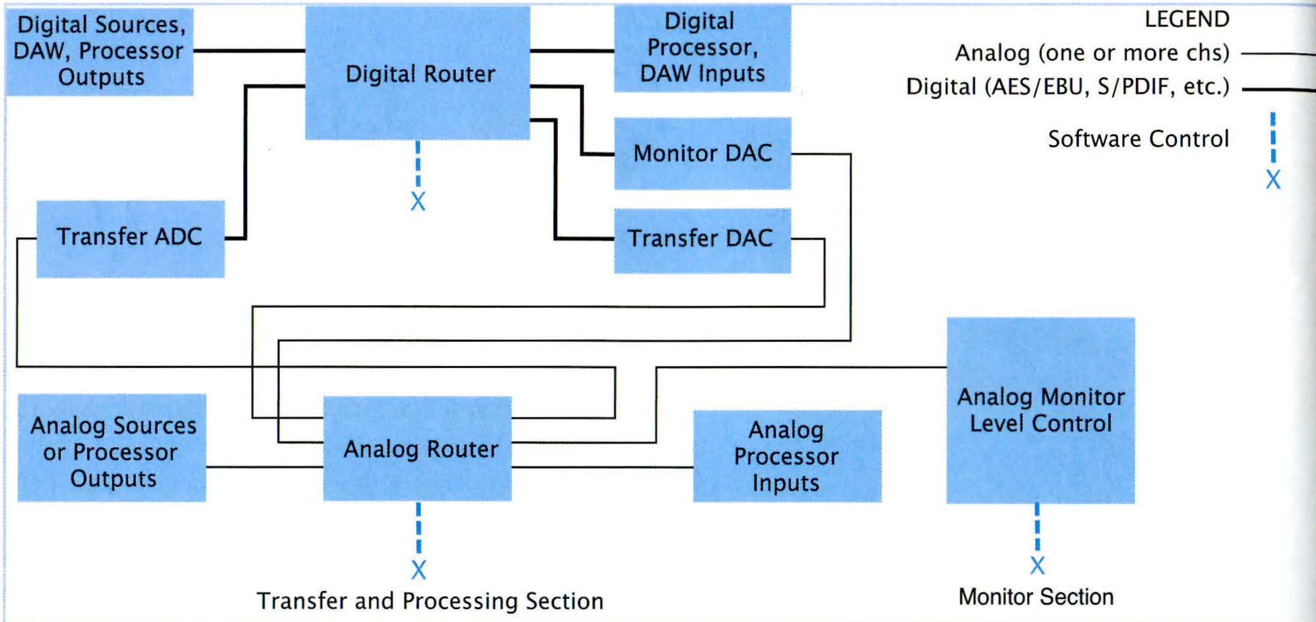
Connecting It All Together

I. Introduction

Unlike mixing studios, mastering studios may change their configuration several times during a busy day. One morning, the mastering engineer might spend an hour auditioning clients' mixes, deciding whether they are ready for mastering or may need some mix revisions; these could be at varying sample rates, wordlengths and formats. Later that morning, he might do a two-hour revision to an existing project which requires a complete repatch of the room setup. And from afternoon through the evening, a full-album mastering with yet another setup. With the increasing emphasis on singles, it's not inconceivable that 4 or 5 separate setups might be needed throughout the day. Thus, the mastering engineer is highly dependent on the efficiency of his workflow routines, and the power of the equipment within those routines to produce consistent and repeatable results.

II. The Modern Mastering Studio

Let's take a close look at the connections in the "ideal" audio mastering studio (pictured next page. The input side of each processor is at its left and output at its right). We'll address the software control aspects of the diagram later in this chapter. Digital sources for routing could be DAW, CD, DVD, Blu-Ray, hard disk recorders, or the outputs of processors such as compressors, limiters, equalizers, reverbs, noise reduction units, etc. Analog sources might be analog tape, LP, the analog outputs of disc players that do not have digital outputs, or analog processors. A digital router is a dedicated digital patchbay that can distribute one source to many destinations and handle multiple



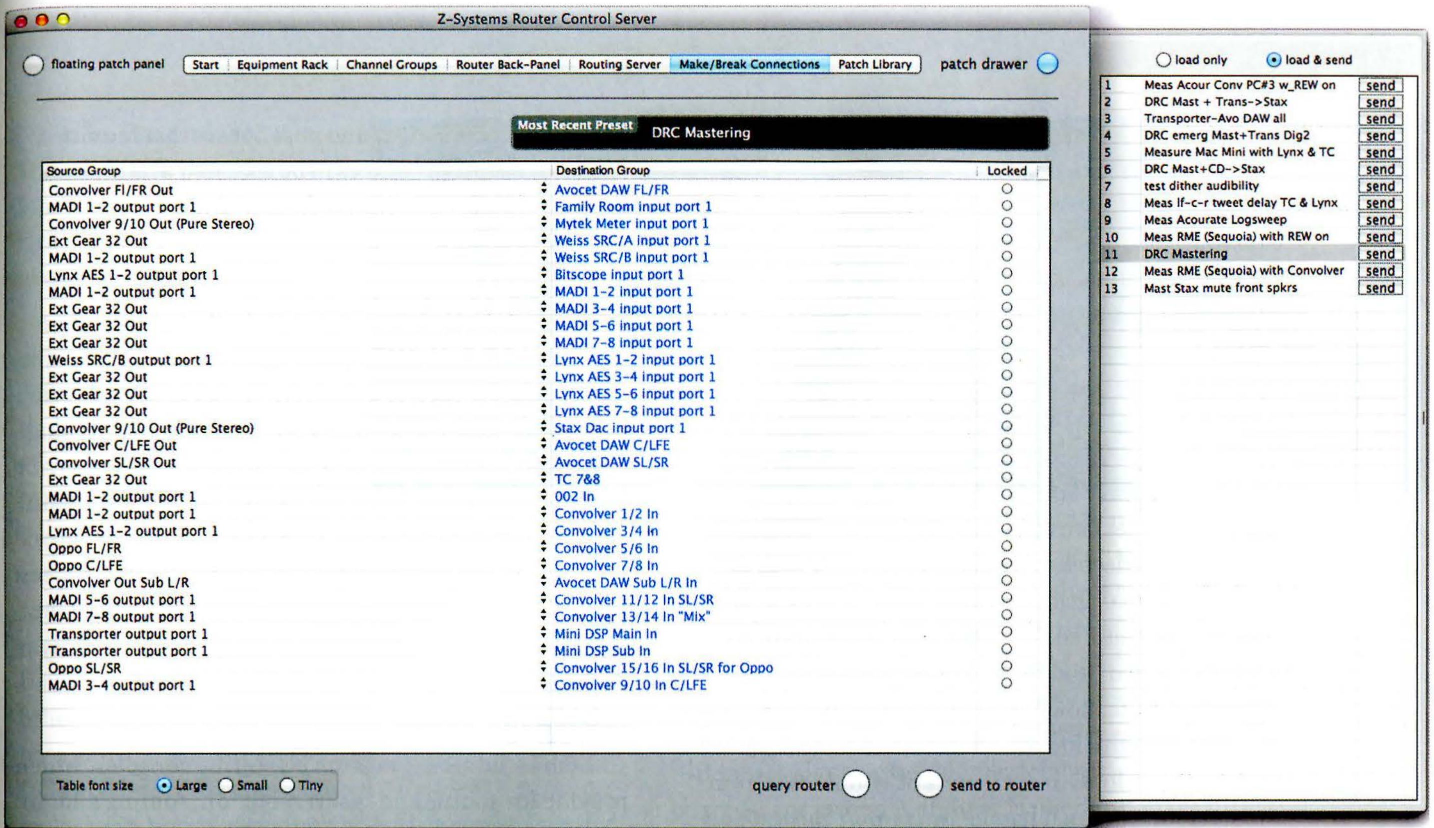
The ideal mastering studio *(block diagram) has all these elements. Routers, manual switches or patchbays interconnect the gear, either manually or under software control.*

impedances—unlike standard plug and jack patchbays. Digital patchbays are inherently more reliable than plugs and jacks and repeatable under software control, so the entire studio configuration can be changed at the push of a button or the click of a mouse. A relay-switched analog router is a patching system. It is also cleaner than plug and jack patchbays, because the relay contacts are sealed from the contamination of the atmosphere, and the contact resistance is lower than that of a jack/plug. Some manufacturers (e.g. Crookwood) perform analog routing under software control. This allows instant reconfiguration and reset of the analog portion of the chain.

Other critical components of the system are: the ADC used to transfer analog tapes to hard disk; the DAC/ADC combination for inserting analog processors in the mastering chain; and the monitor DAC. The latter should be jitter-immune and of the highest-quality (See Chapter 24). Since we are expected to make consistent quality judgments, auditioning all digital sources and pressed media through a single converter guarantees that the monitor gain and sound quality will be consistent. Unfortunately, this principle of consistent monitoring has been subverted by the advent of copy-protected disc media, whose players do not have digital outputs.[1] These make it impossible to proof the final product, or play reference discs through the same converters that were used during the mastering, unless we customize our disc players with digital outputs. This can be accomplished but at some cost. If a single-DAC solution cannot be managed, then the analog output levels of the various DACs have to be matched to ensure consistent monitoring level.

Digital Routing

The digital router connects digital audio sources and destinations in any combination. A single source can be distributed to multiple destinations, but multiple sources cannot be routed to a single destination without



Z-Systems Macintosh Routing Application

a digital mixer. A 16x16 router is sufficient for a small mastering studio, but a medium size studio processing stereo and a small amount of multichannel work would require a 32x32 router, and the largest studios need 128x128 or larger. Since AES/EBU carries 2 channels, a 128x128 router actually switches 256 channels in pairs. There are two basic types of digital routers:

Asynchronous routers, which do not require clocking – such as the Z-Systems Detangler or Crookwood models – can switch virtually any type of signal, support multiple sample rates and different synchronizations in the same chassis, and can be configured for different voltage and impedance standards. Thus, a single unit could be used to route AES/EBU or S/PDIF (2 channels per connection), Dolby E (8 channels per connection), Dolby Digital (6 or more channels), MADI (multiple channels) or other encoded formats, even distribute wordclock and (in some cases) handle composite video – all at the same time! Routing setups can be saved via external software or hardware controllers. Pictured above is the Macintosh-based software application that controls a 32x32 Z-Systems router. It also helps to have a DAC that can switch between sources, which is fundamentally an asynchronous switcher interfaced with a DAC. The Crane Song Avocet (page 194), contains an asynchronous digital switcher feeding up to 8 channels of DAC, combined with a calibrated monitoring facility. Manley and Dangerous Music manufacture economical asynchronous digital switches that can be used to expand a DAC's inputs or do basic routing.

Synchronous routers, which *do* require clocking (either internal or external), are limited to handling only one type of signal (usually AES/EBU or MADI). All signals have to be sampled at the same rate, and framed to the identical clock. Nonetheless, synchronous routers can perform tasks that asynchronous routers cannot do: for example, they can switch signals in the middle of a mastering session without losing clock connections between devices. They can mix multiple sources to a single destination. Good models are bit-transparent, meaning that they make absolutely no changes to the sound quality, and they can carry encoded surround formats to an external decoder.[2] In some routers, such as the RME, individual channels can be split, reversed, polarity reversed, or shuffled. The RME routing software application, called "Totalmix," comes bundled with RME's interfaces. Note, however, that a synchronous router cannot deal with a foreign source – one that is not locked to the system clock, or at a different sample rate from the session. Nor can it handle two sample rates at once. The ideal mastering studio will probably need both an asynchronous and a synchronous router.

Bit Transparency

Digital routers for mastering should be *bit-transparent*: that is, the output should be identical to the source. Asynchronous routers are bit-transparent by design, because they preserve the signal format. But synchronous routers often contain DSP for mixing, panning, etc., so they must be used with caution if a bit-transparent output is required. Avoid routers containing sample rate converters, because they change every incoming signal in order to lock to a common clock. This adds some distortion and is not bit-transparent. Though this type of router is not acceptable in a mastering environment, you will find routers with built-in sample rate converters at broadcast and post houses that require myriads of foreign sources to be accessed at any time and reclocked to a common (*house*) clock.

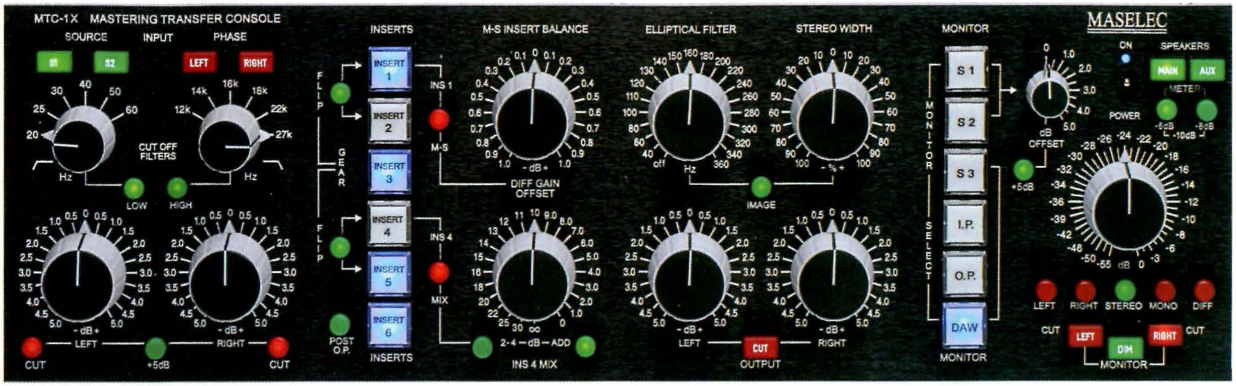
Analog Routing

Pure analog routers are passive switches with no active electronics in the signal path (though of course they do require power to make their connections). Relay-based analog routers have sealed gold contacts, making them much more reliable than a standard plug and jack patchbay. If analog tape is the source, and optimal processing is performed in the analog domain prior to digital conversion, this router should be flexible enough for you to insert analog processors between the tape and the ADC.[3] As shown in the block diagram, digital, analog, or hybrid chains can be created in any desired sequence.

Mastering Console

Some studios use custom mastering consoles, which provide for source and insert selection, routing and, in some cases, processing modules. The implementation, whether via a series of switches, relays, pushbuttons, a patchbay or an analog router, is a matter of ergonomics, sonic quality and personal preference. My own custom analog router can interconnect processors, convert external compressors to parallel and/or MS mode. It also has digital memories, allowing storage and instant comparison of 16 complete setups. For example, I can compare the sound of one compressor and one equalizer with another compressor and equalizer pair, all at the touch of a button. A mastering console could also consist of a set of high-quality analog processors and a few faders in a rack, or laid into a desk surface. Examples of high quality analog mastering consoles available off-the-shelf include the Crookwood M1 Stereo Mastering

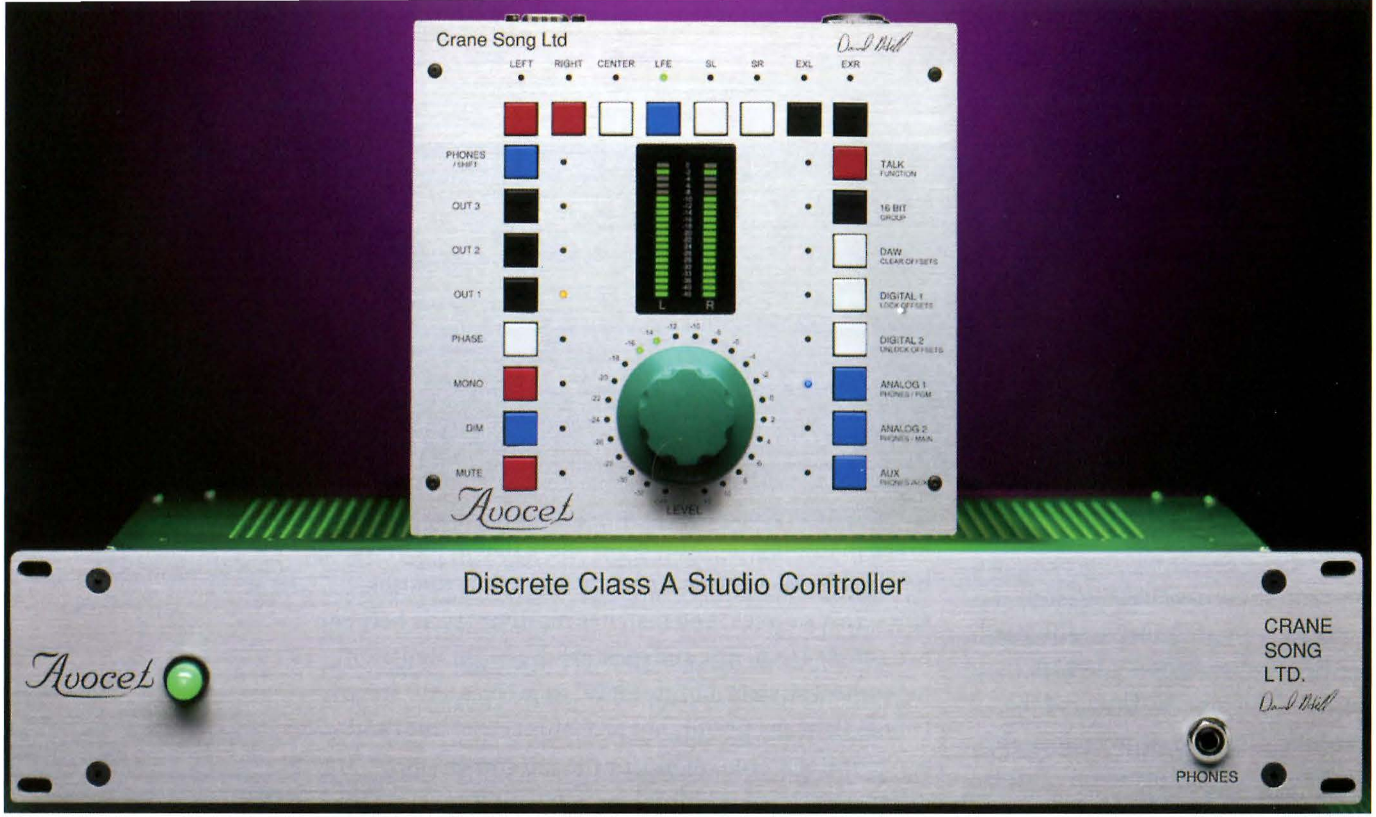
Console, Dangerous Music Master, Manley Backbone, Maselec MTC-1X (pictured at right) and the Sound Performance Lab (SPL) MMC1.



Maselec MTC-1X Mastering Transfer and Monitoring Console

The MTC-1X contains purist, high-headroom analog circuitry. Let's look at some of its features, which every mastering studio needs in some form: The first two-thirds of the faceplate are dedicated to the transfer section, with source processing and routing, and the last third to comprehensive calibrated monitoring. At left are two analog source selectors. Usually, one is the output of the process DAC and the other is an analog tape deck. Next are switchable high pass and low pass filters and polarity inversion. This is followed by 0.5 dB/step channel balance and gain controls. Next, up to six analog processors can be inserted in various permutations. Optional functions include: MS processing, a stereo width control at all frequencies, and an elliptical filter to reduce stereo width below a certain selectable frequency. Elliptical filters were more common for technical reasons in the days of LP, but are suitable for artistic purposes, e.g. combining channels when low frequency information is too separated. A parallel mix chain is provided to allow parallel compression or insertion of a reverb. Output gain/balance of the transfer section is controlled in 0.5 dB steps.

Calibrated Monitoring. I believe that 1 dB/step calibrated monitoring is essential for the 21st century mastering engineer (See Chapter 19), so I am gratified to see many new monitor controllers that meet these needs. The Maselec unit's monitor section chooses among six sources, and matches monitor levels between two sources (e.g. mix and master). Its calibrated monitor control adjusts monitor level in precise 2 dB steps (not as tight as I prefer, but probably acceptable) and can monitor stereo, mono, or the difference signal. The difference signal is helpful for debugging questionable sources. For example, if there is no sound in the difference monitor, then the source must be 100% mono. If there is considerable vocal in the difference monitor, that will be cancelled out when auditioned in mono, which is not very desirable. If there is a lot of reverb in the difference channel, it means that the reverb is either randomly correlated, or the two channels are out of polarity. The latter can be proved by switching to mono to hear if the reverb remains or is cancelled. The defects of AAC processing (space monkeys) are best revealed by listening to the difference signal. Although the MTC-1 is not software controlled, the mastering engineer could log settings by taking a digital photo and putting it into the database, making the job of logging and reset a lot easier than it was before the digital revolution.



Crane Song Avocet Monitor Controller

Monitor Level: DSP or Analog Controlled?

Analog Monitor Controllers. What factors determine the choice between an *analog* or *DSP-based* monitor level control? An analog monitor controller should have high-quality circuitry, be audibly transparent, sound like a "straight wire with gain." It pays to buy from a reputable manufacturer, and also perform the in/out comparisons yourself before committing to the serious investment required for an analog monitor controller or transfer console. Pictured on this page are two high-quality analog monitor controllers with all the features needed to do the right job. The Crane Song Avocet (which we have used in our Studio A) is stereo thru 7.1-capable. The Grace Design M905 is a stereo unit, the M906 is required for surround. Both models

Grace M905 Monitor Controller

have the facility for level-matched comparison between two sources, e.g. the original mix and the master.

Digital Monitor Controllers. The stereo-only TC BMC-2 (pictured at right) makes a great budget monitor controller for a mixing studio, but it does not meet the mastering engineer's need to compare two sources at matched loudness. Nevertheless, I include it here because this is the first unit with quality calibrated monitoring at an affordable price. It has most features mix engineers need (we use it in our mix room). The BMC-2 allows selection of three digital sources, calibrated monitor level, monitor reference, three monitor outputs (analog, digital or phones), metering, and stereo/mono/side selection. The *REF* switch returns the monitor to a known, calibrated gain at any time, and when switched off allows verifying the quality of the sound at lower or higher levels.

A DSP-based monitor controller is suitable if you do not need to listen directly to analog or DSD sources, since the DSD format is incompatible with PCM-based level control. For best sonic performance, a digital monitor level control should have high internal resolution, be dithered to 24-bits, have a high quality low-jitter DAC, and not perform any sample-rate conversion. The main advantage of digital over analog monitor control is cost, provided that the designers have not cut corners on any features that affect sound quality.

I spent a year making the radical decision to change my mastering studio monitor controller from analog to digital. Part of the year was spent comparing and switching between the analog and the digital systems, to confirm that the digital system sounds as transparent, pure, and clean as the analog system. The motivation for change was to enable digital time-alignment of the subwoofers with the mains, since the subwoofers in this room have the most linear frequency response when located in the corners, which are 3 feet farther from the listener than the main speakers. The digital monitor controller includes a digital crossover, transducer linearization and phase correction, and digital equal-

TC Electronic BMC-2 Monitor Controller. Recommended for budget mixing studios and budget or startup mastering studios

Wire #	From	To	Termination	Length	Comments
82A	RME ADI-642 #1 AES 1-2	Z Sys 9 in	X-X		Connected
83	RME ADI-642 #1 AES 5-6	Z Sys 11 in	X-X	3'	Connected
84	RME ADI-642 #1 AES 7-8	Z Sys 12 in	X-X	3'	Connected
85	Weiss SRC/B Out	Z Sys 8 in	X-X	3'	Connected
85A	Lynx (Convolver) 5 Out	Z Sys 1 In	X-X	3'	Connected
84A	RME ADI-642 #4 AES 1-2	Bricasti AES 7-8 In	X-X	3'	
86	Bricasti AES 7-8 Out	RME ADI-642 #4 AES In 1-2	X-X	3'	
87			R-XM	3'	
88	RME ADI-642 #3 WC/T	RME ADI-642 #2 WC/T	B-B	6'	
89	PB B14	TL2	B-B	6'	
90	RME ADI-642 #3 MADI Out	To RME Madi Card In	B-B	20'	
91	From RME Madi Card	RME ADI-642 #1 MADI In	B-B		
92	Z Sys 2 Out	Mini DSP Mains L/R	X-X	5m	Connected
93	RME ADI-642 #1 AES 3-4	Z Sys 10 in	X-X	1,5m	Connected
94	Z Sys 4 Out	Family Room DAC	X-X	5m	Connected
95	Z Sys 23 Out	Avocet DAW FL/FR In	X-X	5m	Moved to #23 out to get 192 kHz
96	Z Sys Router 3	Mini DSP Sub L/R	X-X	5m	Connected
97	Bass Mgr C Out Variable	Subwoofer C In	Captive RCA Male		
98	TC 6000 DB25 1/2 Out	RME ADI-642 #2 AES In 1-2			
99	TC 6000 DB25 3/4 Out	RME ADI-642 #2 AES In 3-4			
100	TC 6000 DB25 5/6 Out	RME ADI-642 #2 AES In 5-6			
101	TC 6000 DB25 7/8 Out	Z Sys 17 In			
102	RME ADI-642 #2 AES 1-2	TC 6000 DB25 1/2 In			
103	RME ADI-642 #2 AES 3-4	TC 6000 DB25 3/4 In			
104	RME ADI-642 #2 AES 5-6	TC 6000 DB25 5/6 In			
105	Z Sys 17 Out	TC 6000 DB25 7/8 In			
106	RME ADI-642 #2 AES 7-8	HEDD AES In	X-X		
107	HEDD AES Out	RME ADI-642 #2 AES In 7-8	X-X		
108	Bass Mgr LFE out variable	Right Subwoofer LFE In	Captive RCA Male		
109	SPARE	(Apogee purple from Z Sys to underneth Avoce	X-X		Spare near Z Sys 18 out
110	Transporter AES out	ZSYS 6 In	x-x		Connected
111	SPARE	Marantz CD Left In	XM-XM	10'	Location: From Marantz to under bot
112	SPARE	Marantz CD Right In	XM-XM	10'	Location: From Marantz to under bot
113	SPARE	(between Z sys and underneth Avocet)	X-X	10'	
114	Z Sys 22	Avocet DAW Sub L/R	X-X	10'	
115	SPARE	(between Z sys and underneth Avocet)	X-X	10'	
116	Spare	Spare between z sys and underneath Avocet)	X-X	10'	was Avocet Dig 1 Sub L/R
117	Spare	Spare between z sys and underneath Avocet)	X-X	10'	was Avocet Dig 2 Sub L/R
118	Spare Rack Snake 2	Avocet Analog 2 RS In	R-XM	10'	
119	Avocet LF Meter out	Dorrough LF Meter In	XF-XM	10'	
120	Avocet RF Meter out	Dorrough RF Meter In	XF-XM	10'	
121	Avocet LS Out	Hafler LS In (via atten)	XF-XM	4'	
122	Avocet RS Out	Hafler RS In (via atten)	XF-XM	4'	
123	Z Sys 20 Out	Avocet DAW In C/LFE	XF-XM	10'	

Studio Wire Number List

ization to linearize the low frequency issues faced by small studios of all sizes. Bass management is included for a complete surround system with stereo subwoofers. Some people question the wisdom of using a digital attenuator in a critical monitor chain; however, from a technical point of view, this monitor controller has an internal resolution of 64-bit floating point, and is dithered to 24 bits, which eliminates all quantization distortion (dither will be explained in Chapter 15). When aligned properly, the digital monitor controller functions exactly like an analog system with equivalent headroom and equal or better noise floor. There are no audible losses, and it has a pure tone quality with no artifacts. There will always be Luddites who are against this idea, but from my point of view this digital controller has proved sonically as transparent as the best analog monitor controllers. It's early in the game, but some of the best-sounding loudspeakers today are now digitally-corrected. So you'll see more mastering engineers accepting digital processing in the monitor chain. I will detail the other advantages of a well-designed digital monitor controller in Chapters 19 and 21.

III. Software Control

The digital router is the only device under software control in most studios, so the analog routing has to be manually configured. But transfer consoles like the MTC-1 are quick and easy to set up. Complex chains of analog and digital components can be created or reconfigured in minutes without plugging in a single patchcord.

I have two digital routers and one analog router. All three have memory recall, so it only takes seconds to completely patch an entire session, making revisions a cinch. Crookwood makes an off-the-shelf software controlled system that can reconfigure all the analog and digital patches at the push of a single button.

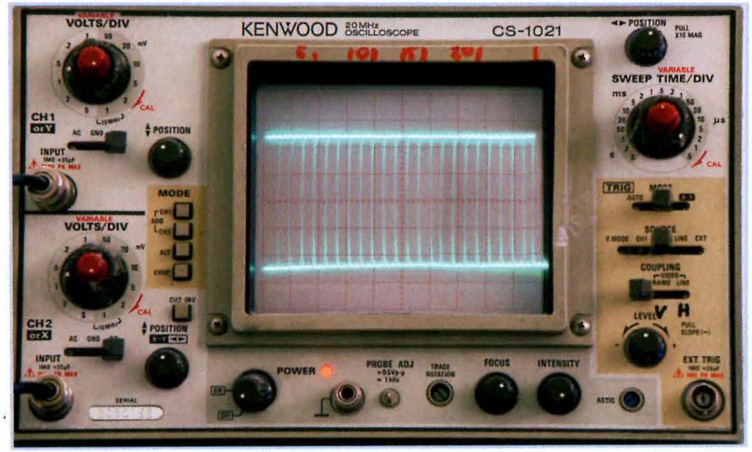
IV. Block Diagram and Wire Numbers

When you construct a mastering studio, it's best to begin with a detailed block diagram, and insert wire numbers from a separate wire number list. Here (page 196) is an example of a wire number list.

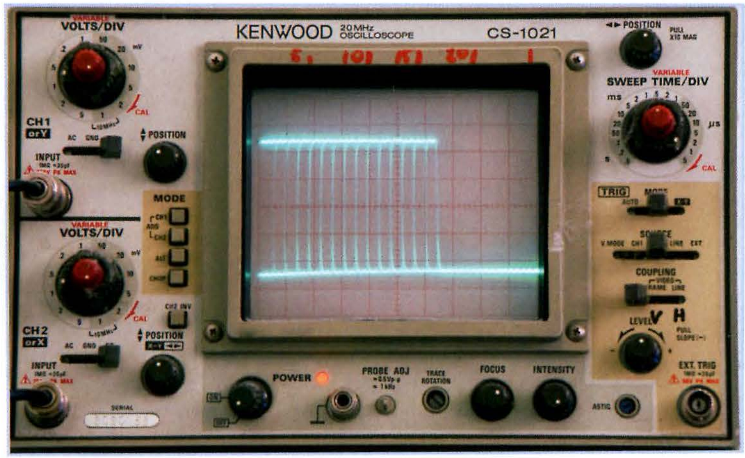
Proper grounding and wire layout techniques are crucial for minimizing signal interference. When the power is distributed from a central location, the best way to avoid ground loops and greatly reduce system noise is to ensure that each power outlet has a home run directly to the distribution, and that the audio follows a similar route to the power.

V. Other Equipment

The bitscope, pictured at right, serves to double-check the bit-integrity of the source. It also confirms that there are no extra bits due to hardware or software bugs, and that the dither appears to be functional (See Chapter 15).



24 Bits active on the bitscope

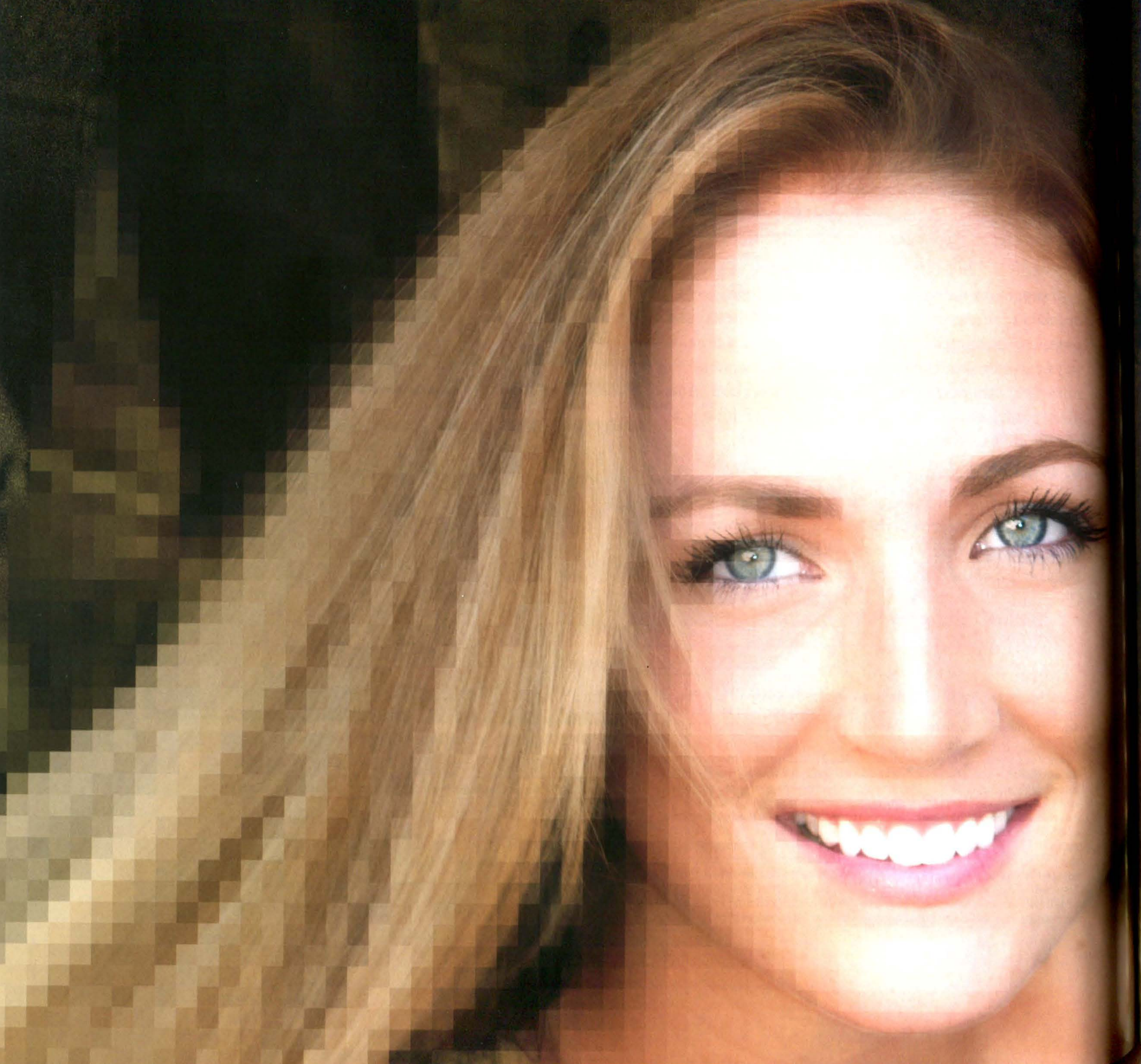


16 Bits active on the bitscope, truncated after the LSB

1
Some disc players have S/PDIF outputs which are limited to 48 kHz. The higher resolution digital outputs are HDMI, which is copy-protected and cannot be accessed by standard DACs.

2
A well-designed digital mixer can be used as a synchronous router, even for encoded sources. But encoded sources (e.g. Dolby Digital) cannot be mixed without losing their coding, so only one-to-one routes are permissible for encoded sources. For non-coded sources, a good digital mixer can mix sources and be bit-transparent to each source, as long as the sum of all the levels does not exceed full scale (0 dBFS). In other words, if you mute all but one of the sources you're mixing, the output of such a mixer is a perfect copy of the input. For a mixer to be bit-transparent, the levels must be set to exactly 0 dB and the pan controls set to produce unity gain. Depending on the brand of mixer, unity pan can be centered or full right or left. Regardless, never turn your back on digital, and test the mixer for bit-transparency (Chapter 22 discusses tests for bit-transparency).

3
To commit or not to commit? That is the question. If you are the type who commits on load-in, I recommend keeping on archive an unprocessed safety transfer of the precious analog tape so another transfer would not be required if a different sound is desired later.



CHAPTER 15

Wordlengths and Dither

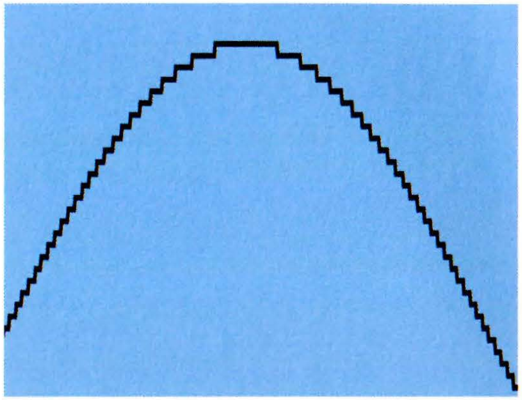
Introduction

Although audio engineers must learn how to deal with and take advantage of wordlength (bit depth) and proper dithering practices, we must also keep our problems in perspective. If the mix isn't good, or the music is not working, then dither probably doesn't matter much at all. But if everything else in a project is right, and we want to maintain the sound quality, then proper dithering is very important.

I. Dither in the Analog Domain

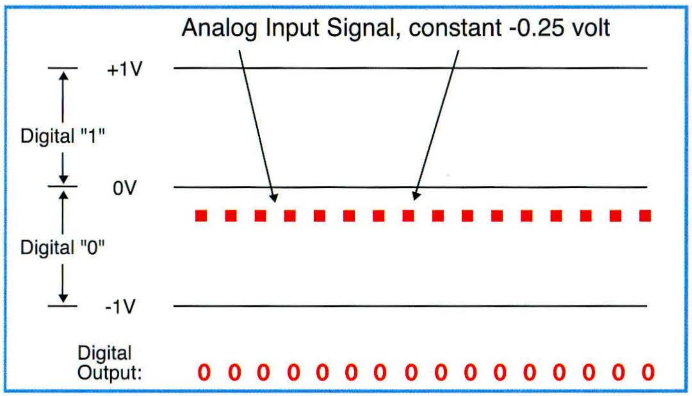
In an analog system, the signal is *continuous,*[1] but in a PCM digital system, the amplitude of the output signal is limited to one of a set of fixed values or numbers. This process is called **Quantization**. Each coded value is a discrete step. For example, there are exactly 65,536 discrete steps, or *values,* available in 16-bit audio, and 16,777,216 discrete steps in 24-bit audio. The approximate codable dynamic range of any PCM system is calculated by multiplying the wordlength by 6: e.g., 8 x 6 = 48 dB for an 8-bit system. So the lowest value that can be coded by 16-bit is 96 dB down from the top; in 24-bit it is 144 dB. If a signal is quantized without dither, this will induce a **distortion** related to the original input signal, introducing any of the following undesirable effects:

- harmonics
- harmonics aliased down to lower frequencies
- intermodulation
- any of a set of highly undesirable kinds of distortion, perceived as a buzz, grit, harshness, coldness and/or loss of depth in the sound.



Section of a sinewave quantized without dither (illustrating "stepping").

Pictured above is a portion of a sine wave that has been quantized without dither.[2] Since there is no resolution below the level of each quantization step, the result is a stepped, distorted waveform. Low-level information is completely lost. To prevent this kind of distortion, we use dither, which is a process that mathematically removes the highly undesirable distortions entirely, and replaces them with a fixed noise level.



Graph of a hypothetical ADC whose LSB threshold is 0 volt. Each sampled analog input is represented by an orange square; in this example, the analog source is held at a continuous -0.25 volt. Note that any input between -1 volt and 0 volt will be lost, because it is below the threshold of the LSB, producing a string of zeros. Because it is below threshold, this signal at -0.25 volts will not be detected, and it will be truncated to a value of 0.

Here's a simple *thought experiment* that explains why dither is necessary and how it works.[3] Let's create a basic ADC. We'll make it sensitive to DC, and bipolar, so it responds to both positive and negative analog inputs. LSB means **least significant bit**, the bit with the smallest (lowest) analog value in the PCM system. We'll give our ADC a very big LSB size of 1 volt to make the numbers easy, and set the LSB threshold at 0 volt. An LSB threshold of 0 volt means that *any signal below 0 volt will be truncated* if dither is not applied. We'll construct our ADC so that an analog source in the range between -1 volt and 0 volt produces a digital output word of 0, and an analog source in the range between 0 volt and +1 volt produces a digital output word of 1. If, without applying any dither, we present a -0.25 volt DC (continuous) signal to the input of the ADC, the output of the ADC will be a string of zeros. Any information below the LSB threshold has been completely lost (pictured below left). Each moment in time to be sampled is represented by an orange square, and each output result by a number. In this case, all the output numbers are zeros.

If we remove the -0.25 volt signal and apply dither to the input of the ADC in the form of a completely random signal (i.e., noise) centered around 0 volt, its peak amplitude will randomly toggle the LSB of the ADC (Figure A, page 201). Now the output of the ADC is a stream of very small random values whose *average is zero volt* (there is an equal number of 0's and 1's).

Leaving the dither on, let's apply our -0.25 volt signal again (Figure B). At each sample point (in time), the -0.25 value of our analog signal is added to the random dither value. The output is now a stream of numbers whose average is equivalent to -0.25 analog volts.[4] We have thus detected and captured information that was previously lost (even though it's been mixed

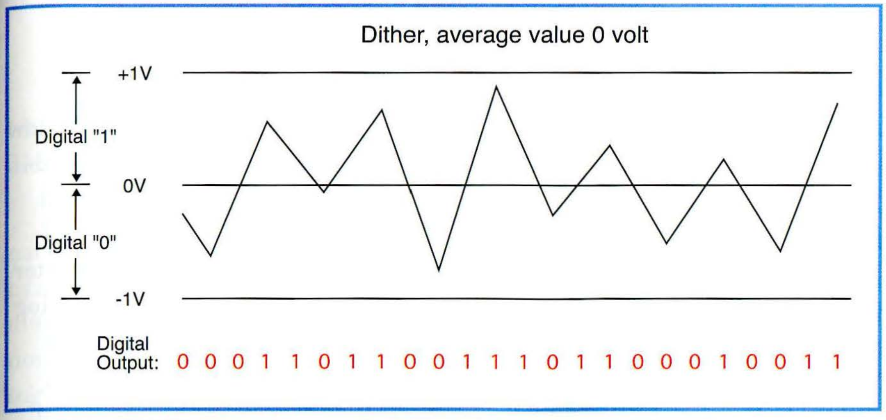


Figure A

Random dither applied to the ADC whose highest peak-to-peak value is slightly greater than the LSB size and whose average value is zero volts.

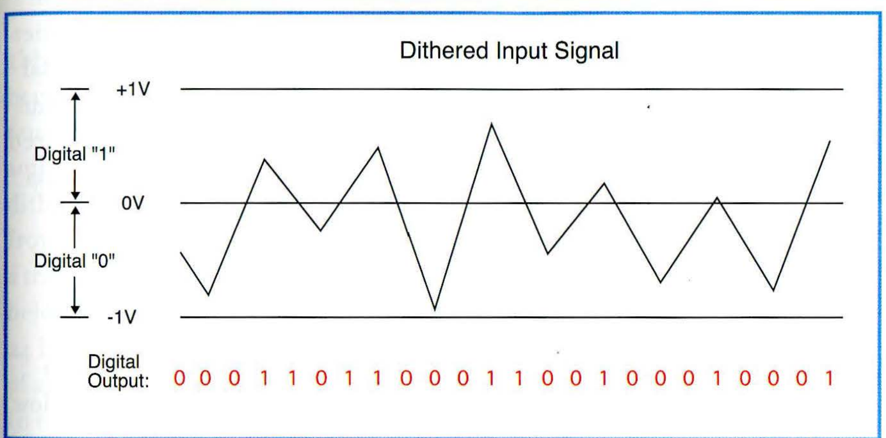


Figure B

Dither summed with the input signal produces an output whose average value is -0.25 volt, the same as the input signal (with added noise). Notice that the input to the ADC looks like the previous dither picture, except it is now biased downward by our -0.25 volt signal.

with added noise). In other words, our low level resolution has improved. The conversion is still essentially random, but the presence of the -0.25 volt signal biases the randomness. Put another way, the characterization of the system with dither on is transformed from being completely deterministic to one of statistical probability. The periodic alternation of the LSB between the states of 0 and 1 results in encoding a source value that is smaller than the LSB. On the average, the LSB puts out a few more zeros than ones because of our -0.25 volt signal. We say that dither exercises, toggles, or modulates the LSB.[5] With the dither on, we can now change the input signal over a continuous range, and the average of the ADC output will track it perfectly.[6]

An input signal of 0.373476 volts will have an average ADC output of (the binary equivalent of) 0.373476. The same will hold true of inputs going over the 1 threshold: an input of 3.22278 will have an average ADC output of 3.22278. So not only has the dither enhanced the resolution of the system to many decimal places, it has also eliminated "stepping".

Dither's resolution enhancement is a true physical/mathematical phenomenon, *not* a means to fool the ear or "to mask the low-level digital breakup." In addition to being able to record and reproduce *all* the analog values at high and medium levels, dither lets us capture low-level signals *below the -96 dB limit for 16-bit!*[7] We can digitally measure undistorted test tones lower

Figure A

1.2 kHz test tone dithered to 16-bits, at the remarkably low level of -130 dBFS (nominally almost 22 bits). This and even lower levels are measurable but inaudible, masked by the dither noise. Demonstrating that 16-bit dithered audio has measurable resolution below the 16th bit. The noise components in each individual bin showing in the -144 dBFS range add up to an RMS total of about -91 dBFS.

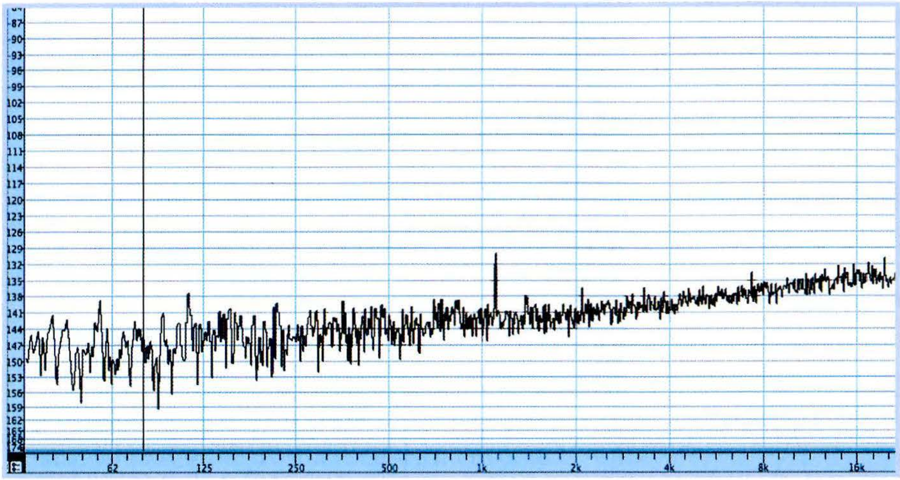
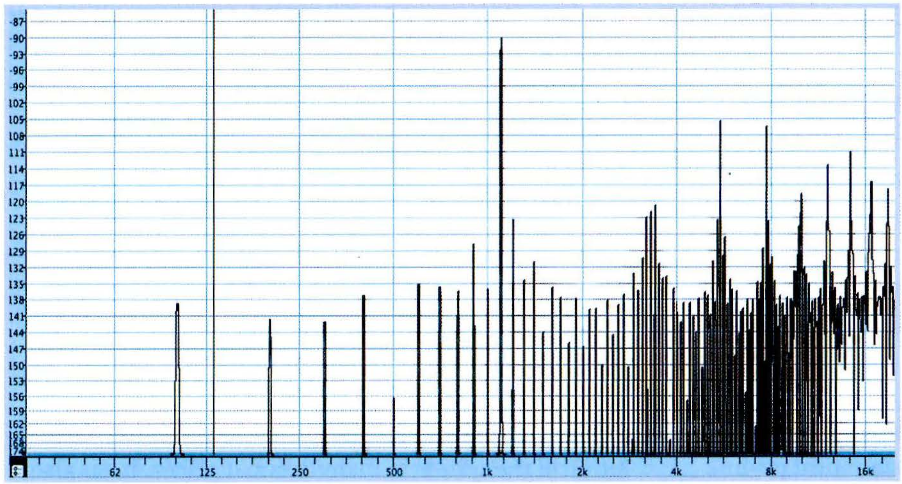


Figure B

Undithered 16-bit test tone at -90 dBFS containing severe audible distortion. Many of these distortion spikes are audible above a 16-bit noise floor so noise added after truncation is ineffective.



than -130 dBFS in a 16-bit dithered system, although -130 is too low to be heard because it is masked by the dither noise (Figure A above). Figure B illustrates the severe distortion that would occur if this signal were not dithered. The perceived (audible) dynamic range of a dithered system is greater than it can code, because human beings are able to hear signals in the presence of noise of greater energy than the signal. We can hear discrete tone signals about as low as -115 dBFS in a 16-bit recording, whose noise floor is about -91 dBFS with the presence of dither.[8] In fact, 16-bit has better low-level resolution than analog tape, and a noise floor quieter than many analog processors. It's a separate issue whether or not we engineers need that quiet a noise floor, or how well digital audio processing serves our musical and sonic desires compared to analog processors and media. We'll leave that discussion for Chapter 22.

What about 24-bit performance? I can hear an undistorted, dithered 24-bit 1 kHz test tone at a remarkably low -140 dBFS through my 24-bit DAC, even though it "only" has a wideband noise floor of about -120 dBFS (equivalent to 20-bit noise). This is about 25 dB better audible resolution than the lowest tone that can be perceived in a dithered 16-bit system. I can also hear the distortion induced when this test tone is not dithered, through this DAC. This shows:

- My DAC has superb low-level signal resolution and decodes all 24 bits of signal
- My room is *very* quiet
- Physics works: A properly dithered system performs

like an analog system. As the signal level decreases, it disappears cleanly into the noise, just like analog.

These results – increased resolution and the elimination of quantization distortion – cannot be achieved by adding noise *after* the A/D conversion. So dither must be added at the proper point in the circuit: adding noise after quantization is no more effective than locking the stable door after the horse has escaped.

With 24-bit conversion and storage, dither is probably not necessary during the original analog encoding, because the inherent thermal noise on their inputs tends to self-dither. Since thermal noise may not be sufficiently random, manufacturers may add their own dither to yield lowest distortion. Likewise, a transfer from analog tape may be noisy enough to self-dither a transfer to 16-bits. But because we can hear signal below the noise floor, I advise encoding analog tapes to 24 bits. There is no advantage in recording the output of any ADC to a 32-bit file, and it uses more storage. In this Chapter the files we discuss are *fixed-point*; in Chapter 16 we will explain *floating-point*.

Testing the gear: The perceived dynamic range of an ADC at any frequency can be measured with an accurate test-tone generator, a good DAC, and a low-noise headphone amplifier with sufficient gain. To conduct the test, simply listen to the analog output while lowering the level of the tone, and find when it disappears (use a high-quality DAC for this test). Another important test for the DAC is to attenuate music in a workstation (about 40 dB) and listen to the output with headphones. Listen for ambience and reverberation: a properly-dithered system with good low-level resolution will reveal ambience, even at that low level.

II. The Need for (re)Dither in the Digital Domain

The First Secret of Digital Audio: How Wordlengths Expand

Let's face the music: as soon as we transform audio by changing its level, equalizing, compressing, or any other calculation that requires multiplication, *its wordlength increases*. For example, after processing, a 16-bit source grows to 32 bits *or more*. This means sound quality will deteriorate if we simply truncate that product to a shorter wordlength. Let's see why this occurs.[9]

Digital audio is all arithmetic, but the accuracy of that arithmetic, and the way the engineer (or the workstation) deals with the arithmetical product can make a meaningful difference to the final sound. All DSPs (Digital Signal Processors) tackle digital audio on a sample-by-sample basis. At 44.1 kHz, there are 44,100 samples in a second. When changing gain, the DSP looks at the first sample, performs a multiplication, produces a new number, then moves on to the next sample. It's that simple.

To simplify our discussion, let's spend some *digital dollars*. Suppose the value of our first digital audio sample is expressed in dollars instead of volts, for example $1.51. And suppose we want to reduce it by 6 dB. 6 dB is half the original value.[10] So, to attenuate our $1.51 sample, we divide it by 2.

This creates a problem: $1.51 divided by 2 equals 75-1/2 cents, or $0.755. So, what should we do with the extra decimal place we've just gained? It turns out that being able to deal effectively with extra places is what good digital audio is all about. If we just drop the extra five, we've lost just half a penny – but in the audio world that half a penny contains a great deal of the natural

MYTH:
"The source was already dithered so I don't need to dither again when I process."

Resolution, Wordlength and Precision

Resolution is an overused term that we must define. Here we use the term **resolution** to indicate whether a source signal of a given level will be represented in the output. This can be expressed as a number of equivalent bits. We define **audible resolution** as the lowest signal which can be heard above the noise (applicable to pure analog systems or hybrid analog/digital systems). We define **measurable resolution** as the lowest signal which can be detected above the noise, e.g. Fig. A page 202.

We define the term **wordlength** (also known as **bit depth**) as the number of discrete data bits employed to transfer a digital value from the source to a destination.[13] **Precision** is defined as the internal data wordlength within the algorithm. A processor's precision must be significantly greater than its output wordlength: e.g. equalization is a complex, multi-step process. It is not easy to determine the internal precision of a processor. We depend on the skill and reputation of the programmer and the sound quality of the result.

ambience, reverberation, decay, warmth, and stereo separation that was present in the original $1.51 sample. Multiplications generally result in a longer wordlength than we started with, and the wordlength can increase, up to the precision of the DSP. For example, a 1 dB gain involves multiplying by 1.122018454 (to 9 place accuracy). $1.51 multiplied by 1.122018454 equals $1.694247866. Although the lower decimal places may seem insignificant, remember that DSPs perform *repeated* precision calculations for filtering, equalization, and compression, so *unless adequate precision is maintained*, the end number may not resemble the right product, yielding distortion. The higher the precision (up to a reasonable limit), the better the resolution and the cleaner the digital audio (see sidebar at left).

Today, DSP-budget is no longer an excuse for not dithering. Even cheap DAWs and processors have enough bandwidth to dither. Inside a digital mixing console (or workstation), the mix bus must be much longer than 24 bits, because adding two (or more) 24-bit samples together, then multiplying by a coefficient (the level of the master fader is one such coefficient), can result in a 48-bit (or larger) sample. The low-level (ambience) information present in the original wordlength is now spread proportionally over a much longer wordlength.

Since the AES/EBU standard can carry up to 24-bits, external processors calculate at the highest possible precision, then bring this long word down to 24 bits and send the result to the outside world, which could be a 24-bit storage device (or another processor). The next processor in line also must reduce its internal long wordlength back to 24 bits for AES/EBU transmission. If not dithered, a slowly cumulating error from process to process adds distortion.

How Dither Works in the Digital Domain

Since truncation is so bad, what about rounding? In our digital dollar example, we ended up with an extra 1/2 cent. In arithmetic class they taught us to round numbers according to the rule "even numbers...round up, odd...round down." But rounding is little better than plain truncation: it still adds lots of quantization distortion. So, when dealing with more numerical precision and small numbers that are sonically meaningful, we still have to use dither to bring the information from the LSBs into the bits we intend to use.

When processing digitally, we add dither in a way similar to the analog approach, except that the processor must generate the dither noise as a series of random numbers. This technique is often called **redithering**, because the signal may have been already dithered during the encoding (recording) process, giving rise to a misconception that additional dither is unnecessary. *When we process digitally, there is no such thing as self-dithering:* No matter how much hiss or noise exists in a source, the noise from the original dither becomes irrelevant. Like the program material, the source noise has now been distributed across a longer wordlength, due to the processing. We must add new dither to preserve resolution before truncation. In the analog example, we learned that the dithered result contains all the low-level information below the LSB, because we *added* the analog dither to the analog signal. Similarly, in the digital domain, we add two digital numbers together, one of which is the digitally-generated dither.

To do this, we calculate random numbers and add a different random number to every sample. Then, cut it off at the destination wordlength. The random numbers must also be different for left and right samples, or else stereo separation will be compromised.

For example, starting with a 24-bit word, to redither it to 16 bits:

```
             ---Upper 16 bits--- -Lower 8-
Orig 24-bit MXXX XXXX XXXX XXXL YYYY YYYY
Add random number                ZZZZ ZZZZ
```

The result of the addition of the Z's with the Y's gets carried over into the new least-significant bit of the 16-bit word (LSB, letter L), and possibly higher bits if we have to carry. Just as in the analog example, the random number sequence combines with the original lower-bit information, modulating the new LSB. The result is that much of the resolution and sound quality of the long word is carried up into the shorter word. Random numbers such as these translate to random noise (hiss) when converted to analog. One generation of 16-bit hiss is so low it is usually inaudible at normal monitor gains.

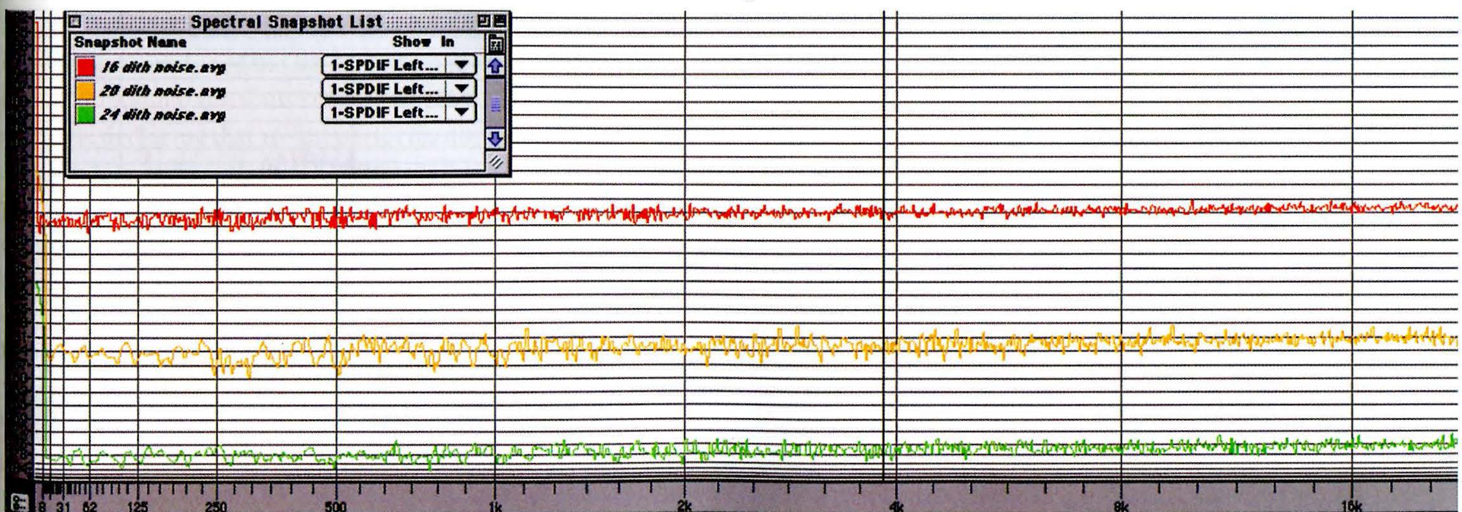
Some Tests for Linearity

To test whether a digital audio workstation truncates digital words or does other nasty things, the only measurement instruments we need are headphones with sufficient gain and a test disc which I produced. Track 42 of *Best of Chesky Classics and Jazz and Audiophile Test Disc, Vol. III*[11] is a fade-to-noise test without dither, demonstrating quantization distortion and loss of resolution. Track 43 is a fade-to-noise with flat dither, and track 44 uses noise-shaped dither (to be explained). Using Track 43 as the test source, it is possible to hear smooth and distortion-free signal down to about -115 dB. Track 44 shows how much better it can sound. If we then process track 43 with digital processing (with and without dither), we can hear what it does to the sound, especially when reduced to 16 bits. If the workstation is not up to par, the result can be quite shocking.

The Effect of Masking

Pictured below, we compare the levels of 16, 20, and 24-bit flat dithered noise. The level of the 16-bit dither (red trace) is about -127 dBFS on this graph, but its wideband level is -91 dBFS. This seems like so little noise, but there is a tradeoff between its benefits (distortion removal and ambience recovery) and the masking effect of the noise. Though the noise is not directly audible at normal listening levels, critical listening demonstrates that sometimes dither noise

MYTH:
Adding noise is the same as dithering.



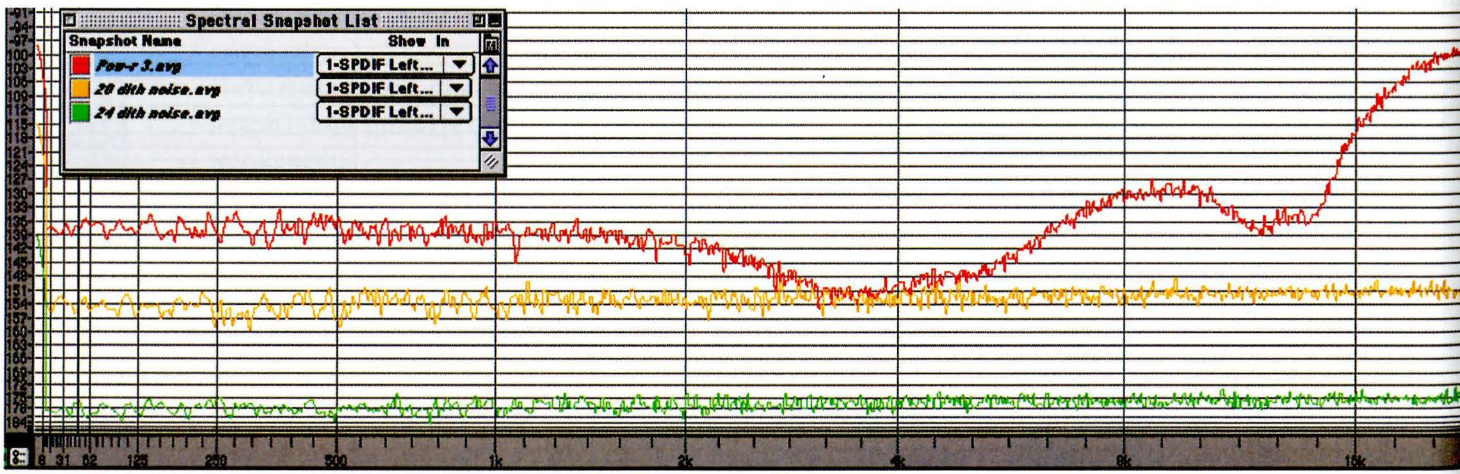
Comparing 16, 20 and 24-bit flat-dithered noise floors (red, orange, green traces, respectively). Though the individual bins of the 16-bit dither measure approximately -127 dBFS, their RMS sum is approximately -91 dBFS.

masks or obscures the very ambience and spatiality we are trying to recover! This is especially true with flat dither at the 16-bit level, which to my ears sometimes adds a slight veil to the sound, narrows the imaging and reduces the depth. It's a tradeoff, because dither's benefits outweigh the losses due to masking.

Improved Dithering Techniques

It's possible to shape (equalize) the dither to minimize this masking effect. Noise-shaping techniques re-equalize the spectrum of the dither while retaining its average power, effectively moving the noise away from the areas where the ear is most sensitive (circa 3 kHz), and into the high-frequency region (10-22 kHz) at a low enough level that for most listeners is inaudible. The best of these noise-shapers yield 19-20 bit performance on a 16-bit CD.

Pictured below is a graph of the amplitude versus frequency (at 44.1 kHz/16-bit) of one of the most successful noise-shaping curves, POW-R dither, type 3 (red trace). For comparison, we can see that it equals the level of 20-bit dither (orange) in the critical 3 kHz range. This is what noise-shaping is all about: dropping the noise where it would be most audible. POW-R 3 uses a very high-order noise-shaping filter, with several dips where human hearing is most sensitive, the inverse of the "F" weighting curve that defines the low-level limit of human hearing. There are numerous noise-shaping redithering processors on the market: some are in hardware form, but most are in plug-ins or built into DAWs. The curve does not directly equalize the material, however, when these were introduced, critical listeners complained that the high-frequency rise of the noise-shaping curves appeared to change the tonality of the sound, adding a bit of brightness. But the rise is not the culprit: my listening tests indicated that the tonality change is caused by masking and unmasking in the midrange. I found that shapes with a little boost around 200 Hz sounded brighter than those with a flat lower midrange, and that the dip in POW-R 3 around 2 kHz seems to unmask low-level material in that range, producing an apparently brighter sound



POW-R type 3 dither at 44.1 kHz/16-bit (red trace), compared with 20-bit flat dither (orange) and 24-bit flat dither (green).

with some program material. With other program material, the result is the opposite: POW-R 3 is perceived as warmer. So masking is clearly the issue, and different shapes are applicable for different recordings.

The choice of which shape to use is a tradeoff among depth, transient response and tonality. To generalize: the higher the noise-shaping, the greater the depth, transient response and sometimes the brighter the sound. The lower the noise-shaping, the more accurate the tonality, but the depth and the transient clarity become lost. For example, a warm-sounding 24-bit master that benefits from depth may translate best to 16-bit using a dither with a higher-order shape. When in doubt, choose a moderate or low shape. As a rule of thumb, remember that highly-processed material sounds worse with a high-order noise shape, but material that has undergone little processing may benefit from a high-order noise shape. Many engineers still prefer flat dither, especially for grungy material where noise can mask and warm up low-level distortion. See ear-training exercise #13 in Chapter 2.

There are now myriads of noise shapers on the market from different vendors. One vendor, Apogee Electronics, produced the *UV-22* system in response to complaints about the sound of earlier noise-shaping systems, declaring that 16-bit performance is just fine. Apogee does not use the word "dither" (their noise is periodic, so they prefer to call it a "signal"), and instead of noise-shaping, the UV-22 adds a carefully calculated noise around 22 kHz, with a slight noise improvement in the midband compared to flat 16-bit dither. I spend no more than 5 minutes, usually less, picking the best 16-bit dither for the master I am working with, and note the dither choice in the mastering log. With the decline of the compact disc and the advent of downloads, there is less emphasis on 16-bit dither and more on getting 24-bit material to the distributor for conversion to coded (e.g., Mastered for iTunes). This is much for the better, because I find 24-bit sounds purer and deeper, and translates best to the release medium.

{ "As a general rule: highly processed material sounds worse with a high order noise shape." }

We can effectively compare the sound and resolution of these redithering techniques by first lowering the music level going into the dither unit by about 40 dB, then listening to the output of a high-quality DAC on headphones. The sonic differences between high- and low-quality dithering systems can be shocking: Some will be grainy, some noisy, and some distorted, indicating improper dithering or poor calculation. When making judgments at high gain, try to discount any obvious high-frequency noise due to noise-shaping, because noise-shaping is designed to be inaudible at normal gain.

The Cost of Cumulative Dithering at 16 bits

As we have already seen, the measured amplitude of 16-bit dither is an extremely low -91 dBFS. But a skilled listener does not have to play material loudly to notice the degradation of truncation or improper dithering. At 16 bits, dithering always sounds better than truncation, because inharmonic distortion is very unmusical. But to avoid a potential sonic veil—*let there be only one generation of 16-bit dither in a CD project, the one-time, final process*.[12] Mix to a long wordlength medium and send that file to the mastering house, which will apply 16-bit dither once, at the last stage. Noise-shaping is fragile: 16-bit noise-shaped material should not be further processed, and cumulative noise-shaped 16-bit dithers can sound thin or edgy.

{ *"Avoid the Slow Death: a gradual loss of depth."* }

Diminishing Returns with 24-bit Chains

The effect of cumulative dither noise at 24 bits is so low it should not be a concern; a single dithered digital processor produces -139 dBFS noise, about 20 dB below that of a quiet converter. Even six 24-bit dithered processors in a row raise the noise floor to -131, which is more than 11 dB below the noise floor of a quiet converter. In fact, the issue with 24-bit is not cumulative noise, but whether or not truncation has occurred, which causes distortion *perceivable above the noise*. Within a native DAW, plug-ins communicate at 32 bits or longer, so there is an advantage to producing a 32-bit file. When feeding external digital processors via AES/EBU (which is limited to 24 bits), the 32-bit word should ideally be dithered to 24 bits. And each processor should redither to 24 bits before feeding the next processor in a chain. But not all external processors provide dithering. Is redithering to 24 bits that important? I hear some loss every time a signal is truncated to 24 bits, but it doesn't hurt to be practical, as the losses are usually subtle. So if you like the sound of a processor that does not dither to 24 bits, use it. Just try to avoid 24-bit truncations wherever possible, to avoid the slow death: a gradual loss of depth.

In summary: If your DAW is 32-bit, mix if possible to a 32-bit file, which retains its resolution, or dither to 24 bits, but don't lose any sleep if you have to truncate to 24 bits once in a while.

III. Examples Examples Examples!

What to Do

1) When reducing wordlength, add dither. Example: From a 24-bit processor to a 16-bit CD.

2) Avoid dithering to 16 bits more than once on any project. Example: Use 24-bit or 32-bit intermediate storage, and do not store intermediate products as 16-bit.

3) Wordlength increases with almost any DSP calculation. Example: The outputs of digital consoles, DAWs and processors will be 24- or 32- bit, even if you start with a 16-bit source.

4) *In any project, sample-rate conversion should be the next-to-the-last operation, and dithering to the shortest wordlength must be last.* Intermediate dithering may occur "behind the scenes", e.g., from 48 to 24 within the processors. It's helpful to output a 32-bit file from the SRC, and then feed the dithering: this preserves the most resolution. While truncation should be avoided, if the SRC does not dither internally, truncation to 24 bits sounds far less bothersome to the ear than to 16 bits.

5) High level peak-limited material may overload with the addition of dither; peaks could take it a smidgen past the top, so to be safe, drop the gain of peak-limited material 0.1 dB when dithering or set a limiter ceiling of -0.1 dB.

6) Most software adds the dither and produces the shorter-length file at the same time. But sometimes that has to be done in two steps. For example, invoke the dither (usually) in a plug-in, then bounce to a new file and tell it explicitly to make the new file at the desired reduced wordlength.

7) Every "flavor" of 16-bit dither and noise-shaping type sounds different, and none sounds as good as the 24-bit original (though some come very close). It is useful to choose the "flavor" of dither more appropriate for a given type of music.

8) Since most DAWs are internally floating point (to be explained in Chapter 16), all outputs and busses must be dithered to a fixed point value. When bouncing tracks to *fixed-point* files, if possible, dither the bus to the wordlength of the capture. E.G., if bouncing to 24-bit tracks, insert a 24-bit dithering plug-in on the bus that's feeding that track.
9) When converting from PCM to a lossy-coded medium such as mp3 or AAC, *don't* dither the long wordlength source down to 16-bit. Instead, use the full wordlength (e.g. 24-bits or 32-bit float) to get the most from the conversion, because lossy coding performs better than 16 bits in some frequency ranges.
10) Invoking dither on *input* is a bad idea. Capture the full wordlength of what's coming into a processor, and save any dithering for the final *output* stage.
11) It's amazing how few people know this simple fact: All DACs can only use the top 24 fixed bits of the source, so it's a lossy process to feed them a floating-point source. It's not nice to fool Mother Nature: many DAW manufacturers do not provide the ability to dither the output of their DAW for monitoring, which is absolutely necessary. This explains at least some of the complaints about certain DAWs sounding "smaller." If the monitoring is not dithered, invoke a 24-bit dithering plug-in on any DAW output that feeds a DAC.

What Not to Do

After creating the 16-bit CD master, do not perform any calculations, do not fade out, do not change gain, or you will lose all the hard-earned work you put into creating that file. This is why mastering-grade DAWs permit working with 24- or 32-bit files right up to the point of cutting the CD master. Dithering is applied only at the last step, while the DAW makes the master.

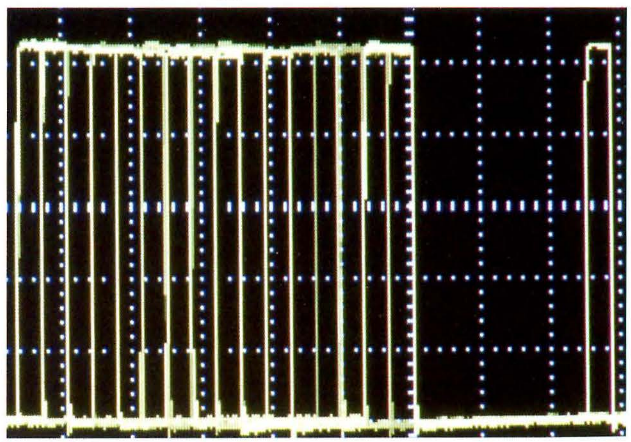
The Pro Tools HD 48-bit Mixer

For those using Pro Tools HD with the HD cards, don't be so quick to jump on the new version, because HD has excellent sonic and technical performance. This mixer has had time to mature; absolutely use the dithered mixer, which is located in the "Plug-ins Unused" folder. Drag it into the plug-ins folder and restart Pro Tools. This mixer reduces an internal 48-bit wordlength calculation to 24 bits. Pictured below is a hardware bitscope photo of a 16-bit file passed through the mixer set to unity gain, showing the addition of 24 bit dither noise plus the signature of the original 16-bit file — an indication the mixer is clean and bit-transparent except for the dither. Be aware that Pro Tools version 11 does not dither its mixer and they expect you to capture long wordlength or add your own dither.

MYTH: *You can't mix wordlengths in a single workstation session.*

SRC and Dither

Think of SRC as another process that expands the incoming wordlength to its internal resolution. Its output wordlength will be as long as the calculation resolution of the SRC, which is probably 32-bit float. This means that sample-rate converting a 1644 source to a 1648 destination is a lossy process! If you must do that (e.g., for a Quicktime video), first convert to 3248, then dither down to 16. The Weiss Saracon SRC can do that in one operation; consider storing the 3248 as a backup. Fortunately, dither noise at 48 kHz is spread around


Pro Tools HD 24-bit dithered mixer at unity gain, fed a 16-bit source signal.

a wider bandwidth, so the cumulative audible noise will not be as egregious as double 1644 dither.

iTunes CD or AAC to WAV

CD Import and Process: Let's say that you want to get the most from a pressed CD by doing some of your own processing. In that case, note that consumer programs purporting to let consumers equalize and process CDs to produce new CDs are really *truncating the data*. I hope you've noticed the loss! I advise against using iTunes for any processing because it is not capable of outputting to a 24-bit or 32-bit file or dithering. In other words, if you start with a 16-bit CD and want to process it while preserving every bit of its purity of tone and depth, then use a professional DAW. Import the CD as a 16-bit file, and export it (processed and dithered) to a 24-bit file. This produces a result with all the resolution of the source. If you have to make a 16-bit result, dither it; it will likely sound a little smaller than the source CD, as this processed copy includes cumulative 16-bit dither noise.

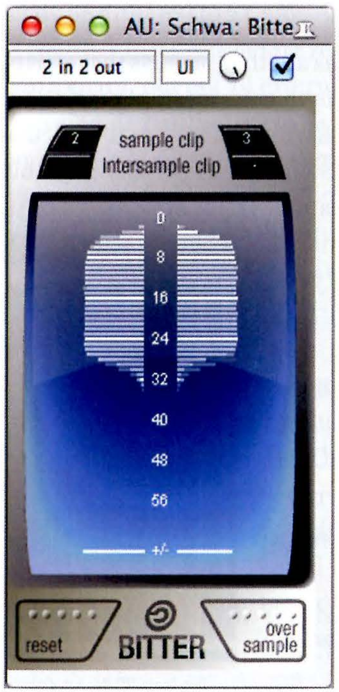
Bitter as a diagnostic tool: A great tool for diagnosing problems or just keeping an eye on things is a free plug-in for Mac or PC from Stillwell audio called *Bitter* (pictured below). Bitter counts bits up to 64-bit float, and also displays sample clips and intersample clips.

AAC Import and convert to WAV: Let's not give AAC a worse name than it deserves: very often we play it back incorrectly. AAC files have an effective psychoacoustic resolution of about 18 bits. So, even if you simply want to import an AAC file and convert it to WAV, it will lose some resolution if you make a 16-bit WAV. Use a DAW that can import AAC files, and import them as 3244 floating point files, which is the output format of an AAC decoder. Then process if desired, or simply dither down to 24 bits for playback on any decent DAC. That's one reason why 32-bit float iTunes doesn't sound as good as dithered DAWs when playing CDs at other than unity gain or with the equalizer turned on, or AACs played at any gain. Screenshots (pages 211 and 212) illustrate the situation.

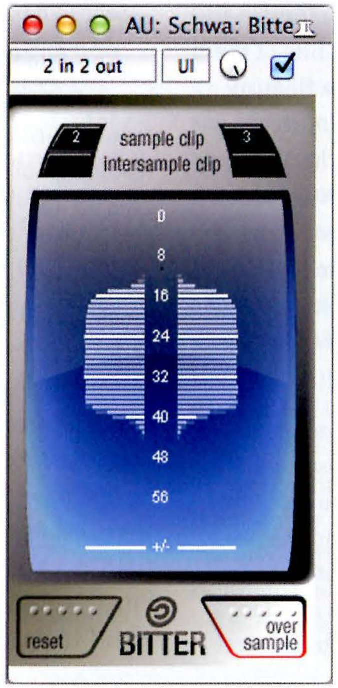
IV. Managing DAW Wordlengths

In older versions of Pro Tools, it was necessary to limit a session to one wordlength. This led to user misconceptions

16-bit wav file played at 0 dB.



Same file, fader reduced slightly. Some sections of the file had clipped and the clip indicator was not cleared.



Same file, with fader reduced 60 dB. Notice that the wordlength grows to over 40 bits.

when bouncing. For example, when bouncing to a new file in Pro Tools, even if the session is 16-bit, the output file should be 24-bit, or you'll truncate data.

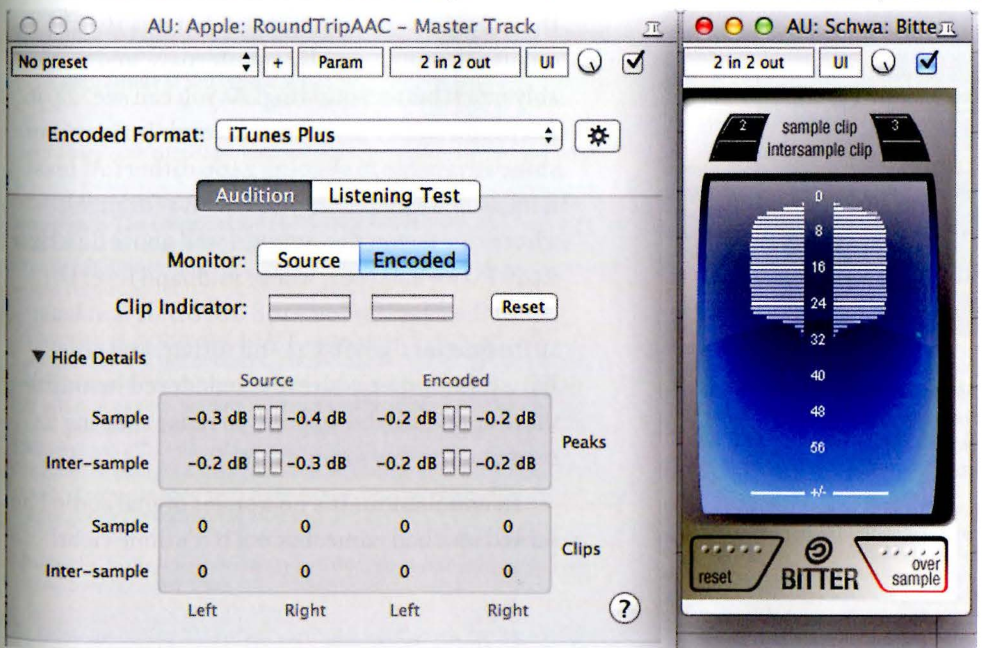
It is not necessary to "expand" the wordlength of the session before mixing: the sound can never get more resolved than what was originally encoded. Regardless of the source sample's wordlength, a workstation will always calculate to its internal precision. You can think of it as adding zeros to the tail of any shorter words: this does not change the original value. In other words, 16, 24 and 32-bit samples can coexist in a well-designed workstation, and when calculations take place, all sample wordlengths are increased to the internal wordlength of the workstation. Plug-ins should be captured to their full wordlength. However, in older versions of Pro Tools and Digital Performer, in order to insert bounced tracks back in the session at the best resolution, it is necessary to convert the session to a longer wordlength. This is only because of the practical limitations of the software, for there is no mathematical necessity. This is an inconvenience to mastering engineers, who regularly mix wordlengths (and file formats) in the same session.

Auto-Dither

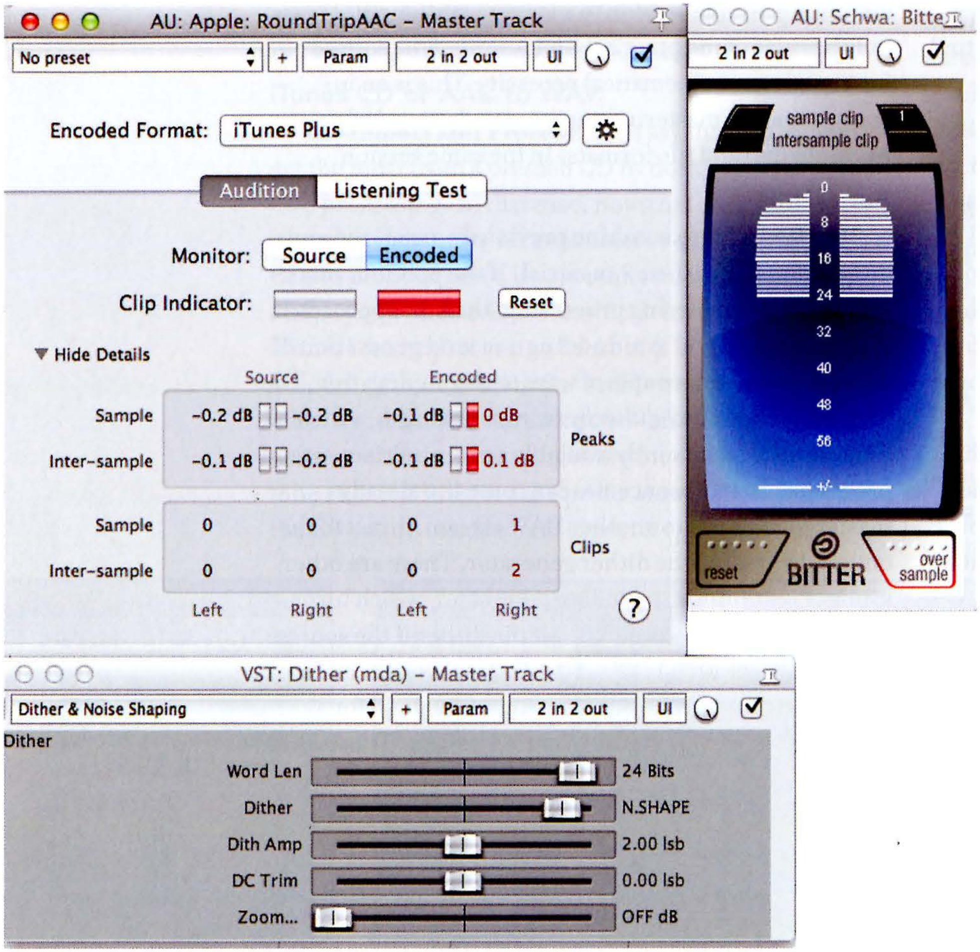
We often have to combine previously-mastered and dithered music with new material. If the previous material needs no mastering processing, the best approach is to try to clone it and avoid adding a second generation of dither. There are a couple of ways to accomplish this. The first is by using *auto-dither by source wordlength*, which is a clever feature currently available only in a Prism-brand processor. In its absence, we can route the already-mastered material to another DAW stream, direct to the output, bypassing the dither generator. There are other kinds of *auto-dither*, including *auto-black*, which turns off the dither if the source audio level goes below a certain threshold for a period of time. This is useful if the producer insists on total silence between pieces.

V. Dither At High Sample Rates

Moving to high sample rates automatically provides a signal-to-noise advantage, because 16 bits at 96 kHz is 3.4 dB quieter in the audible band than at 44.1. Noise-shaping at high sample rates can



PCM fixed-point source file encoded to AAC, then decoded to PCM. Notice there are now 32 active bits.

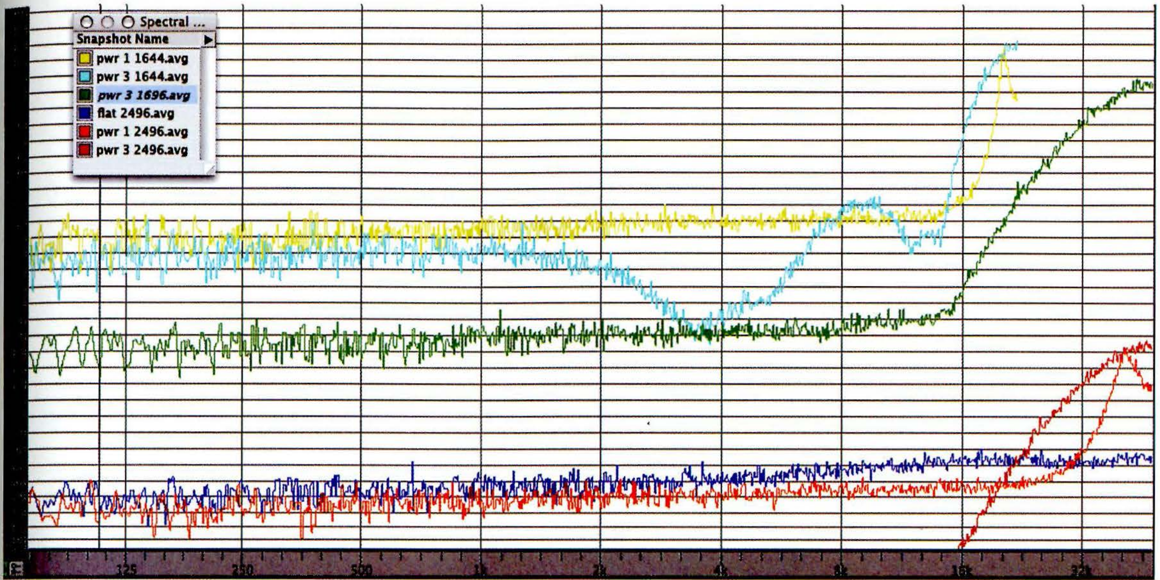


Dither the result to 24 bits so it sounds best when listening on a DAC, or for conversion to a 24-bit wav. Engineers might want to reduce the amplitude to prevent clipping, as shown both in Apple's plug-in and in Bitter. We don't hear the 24-bit dither noise directly — its amplitude is too low, but we do notice the slight improvement in sound due to the reduction of distortion, which would have been above the noise floor of the DAC.

allow shorter wordlength files with a very low psychoacoustic noise floor. The noise can be made extremely low and flat in the audible band and the shaping moved above 20 kHz. In fact, 1696 noise-shaped can sound nearly as good as 2444, as I discovered one day when I accidentally left 16-bit dither on while working at 96 kHz.

This graph (page 213) compares dithers at different sample rates, from highest to lowest level, demonstrating how designers attempt to squeeze 24 bits of information into a 16 bit container. 1644 POW-R 1 (yellow) is very flat with an extreme rise above 20 kHz. 1644 POW-R 3 (turquoise), with its extreme shaping. Moving to a higher sample rate, 1696 POW-R 3 (green) is as low as flat 2044 dither in the audible frequency range — its rise is postponed to almost 16 kHz. In the midband it is totally flat, but it is 16 dB quieter than its 44.1 kHz counterpart, and probably much better sounding! As you can see, 2496 flat dither (blue) is very low in level. Is there any sonic advantage to shaping 2496 dither? At least it measures quieter: 2496 POW-R 1 (orange), whose rise doesn't occur until well above 32 kHz. 2496 POW-R 3 (red), whose midband level is below the chart (below -198 dBFS!). It is at least 18 dB quieter than flat 24-bit dither, and since flat 24-bit dither is already considered inaudible, there appears to be no point to noise shaping 24-bit dithers at any sample rate!

In conclusion: It's no wonder digital audio has picked up a bad name, but not if it's done right!



Dithers at different sample rates, from highest to lowest level. At 16-bit/44.1 kHz: POW-R 1 (yellow) and POW-R 3 (turquoise). At 16-bit/96 kHz: POW-R 3 (green). At 24-bit/96 kHz: flat dither (blue), POW-R 1 (orange) and POW-R 3 (red).

1
Continuous does *not* mean that the analog signal has infinite resolution. The finite resolution of analog is defined as the lowest signal not covered by the noise floor.

2 Image courtesy of Jim Johnston.

3
Based on original concept by Mithat Konar, director of engineering, Biró technology, with contributions from Robin Reumers and Jim Johnston.

4
As an exercise, count the number of 0's and 1's in this image and take the average. The average of the 1's should result in 9/24, which is 0.375. This relates to -0.25 volt (0.375*2 - 1) on the graph. We only present 24 samples, but since dither is a probabilistic system, an exact measurement would require an infinite number of samples!

5
In practice, more than just the LSB is exercised. It can be all the bits. In base 10, if we add two numbers, and the sum is greater than 9, we have to carry. In base 2, we also have to carry and if the next significant digit to the left is not a zero, we have to keep on carrying until the next digit up is a zero, and turn it into a 1. In 2's complement, the addition of dither at the LSB level will affect the values of many digits, including the MSB, because the number changes polarity between negative and positive. Seen on a bit-scope, it seems to show two values at once, because the numbers are always toggling with the addition of dither.

6
"Perfectly" to the lowest decimal place that can still be accurately determined over the noise floor.

7
Or below the coding floor of any particular wordlength. In other words, if we dither to 20 bits, whose coded range is 120 dB, we can include low level signals below the -120 dB limit, and so on.

8
The noise floor is raised 4.77 dB, to be exact. This is the least amount of noise necessary to properly dither a digital audio signal and eliminate all possible distortion. The statistical distribution of the noise must be triangular probability. See Wannamaker, R. A., & Lipshitz, S. P., & Vanderkooy, J. (1992) Quantization and Dither: A theoretical Survey. *J.Audio Eng. Soc.*, Vol 40, No 7, pp.601.

9
Some processor and DAW manufacturers still have not recognized the importance of internal processor precision, and the fact that wordlengths expand — a prime reason why some digital devices sound sweeter than others. But many have gotten the message and there are many more decent procsssors around.

10
For signals which are correlated, the formula is: dB change = 20 * log (ratio). For example, if we drop the level by a ratio of 1/2, whose log is -.3010, then multiply by 20, the approximate result is -6 dB (6 dB down), to the nearest decibel. Note the use of the word approximate. And yes, the degree of accuracy used in such calculations affects the quality of our audio.

11
Chesky JD111 (I produced this disc). The hard-to-find CBS CD-1, track 20, also contains a fade-to-noise test.

12
Many argue that several generations of 16-bit dither circa -91 dBFS should be insignificant. It depends on the material. Pristine, transparent digitally-recorded material can lose some depth with even one generation of 16-bit dither noise. But some music that depends on distortion (e.g., rock) sounds better with a high noise floor, or with flat dither instead of noise-shaped dither. The psychoacoustic argument goes on, which is why we have ears to make judgments!

13
Thanks to Paul Frindle for this most concise definition of wordlength.

CHAPTER 16

Decibels: Going Deep

Introduction

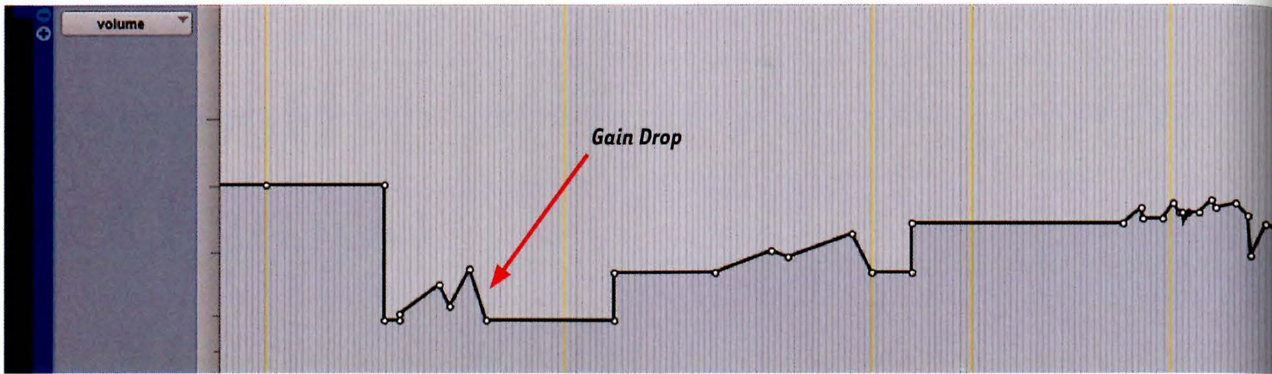
Now that we've mastered wordlength management, in this chapter we will examine decibels in depth: levels, clipping, distortion, signal-to-noise ratio, loudness, loudness normalization, fixed versus floating-point formats, AAC conversion, and more.

I. Stamp Out Slippery Language

An Essay on the V-Word

Have you noticed that I've managed to get through 15 chapters without once using the "v-word?" When a student asks me "how do you measure volume?" I reply by showing him this picture.

This is how we measure Volume



Pro Tools' so-called "volume line" is technically a gain line. This is a graph of a process, not a result. In the marked place, where the gain drops, can you tell if it dropped because the music was getting too loud, or because the producer wanted to purposely create a decrescendo?

You won't find the v-word in physics or acoustics textbooks. In fact, the term **volume** is ambiguous audio parlance. We commonly use it to mean three completely different things: **gain**, **level** and **loudness**. A single meaning for *volume* has not been standardized, so listeners must guess the user's intention from context. The ambiguity causes confusion among consumers and professionals. Let me demonstrate with some examples.

Confusing a process with a result: Pro Tools has been calling its "gain rubberband" the *volume line* forever, so it's become accepted language by now, but the v-word does invite confusing a *process* with its *result*. For example, in the above image, looking at the dotted line, we may ask, "What is the volume now?" (using the word *volume* to mean *output level)*. The correct answer is, "We cannot tell from the information provided." In the marked place, where the gain drops, can we tell by looking at the so-called *volume line* if the change is because the music was getting too loud, or because the producer purposely wanted to create a decrescendo? To answer, we must know both the incoming *level* and the *gain* that's being applied.

The term *volume* has been confusingly used for process and result within the same conversation, giving rise to bizarre dialogues between engineers and our clients that sound like a comedy sketch: "Turn up the volume, please, so I can have more volume." Pro Tools 10 and up have introduced the very welcome term *clip gain*, since in audio language *gain* means one and only one thing: the difference between an input level and an output level. To be consistent, they should rename the other process the *gain line*, because changing gain has always been its task. SADiE has always called their automation line a *gain line*.

Another confusion between process and result arises from calibrated monitor controllers that are marked in SPL, e.g., 85 dB. This is a deceptive practice. The SPL number, intended to be the result, can only be correct under one condition: if the music being played has the same program level as the calibration level. For example, a hot master could play at 86 or even 90 dB SPL while the control is still at the 83 position. We should mark monitor controls like every other gain control: with 0 dB as unity gain. The 0 dB position is calibrated to produce the desired SPL with a test signal. We'll discuss calibrated monitor controls in detail in Chapter 19.

Our clients get confused all the time between process and result. When auditioning a master, they see that their volume control is turned way down, so they think the sound (the "volume") will be low. But the *result* could be loud or soft, depending on the level of the incoming signal. Think of a monitor level control as a water faucet. When the water pressure is low, we have to open the faucet, and vice versa, to get the same output pressure (*pressure* is a measurement of *level*). There is no way to tell what the output pressure will be without knowing both the incoming pressure and the position of the faucet – exactly what happens in digital and analog audio practice.

This is a volume control!

To escape ambiguity, I will avoid using the v-word for the rest of this book. Won't that be fun!

Important Audio Terms and Techniques

Attenuation... when expressed in dB is an optional term for negative gain, e.g., a loss. Example: 20 dB *attenuation* is the same as -20 dB gain.

Decibel... Abbreviated *dB*, is a relative quantity; it is always expressed as a ratio, compared to a *reference*. For example, what if every length had to be compared to one centimeter? You'd say, "this piece of string is five times longer than one centimeter." It's the same with decibels, though sometimes the reference is just implied. +10 dB means "10 dB more than my reference, which I defined as 0 dB." Decibels are logarithmic ratios, so if we mean "twice as much voltage," we say "6 dB more" [20 * log (2) = 6].

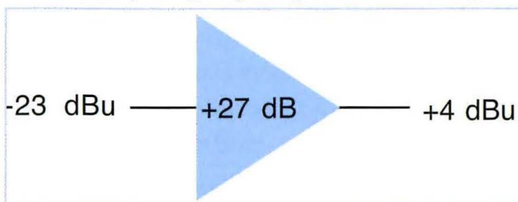
dBu, dBm, dB SPL, dBFS... are ratios with predefined references, so they can be converted to absolute values in volts or power, etc. I believe the term **dBu** was coined in the 1960s; it means *decibels unterminated* (or *unloaded), compared to a voltage reference of 0.775 volts.* **dBm** means *decibels compared to a power reference of one milliwatt.* **dBFS** means *decibels compared to full scale PCM;* 0 dBFS represents the highest digital level we can encode, but as you will see later in the chapter, not the highest that we have to deal with.

Plurals... We *do* say "two decibels", but we do *not* pluralize the abbreviation. We do say "two dee bee," but we do not say "two dee bees."

Gain or **Amplification**... is a *relative* term expressed in decibels with no suffix: it is the ratio of the amplifier's output level to the input level. The term *monitor gain* is so slippery that I prefer using a clear term: **Monitor Control Position.** E.G., we can say "the monitor control is at the 0 dB position."

Intensity... is a measure of energy flow per unit area. For practical purposes, sound intensity is the same as SPL (see below).[1,2]

Level... is a measure of intensity, but when used alone, because it can mean almost anything, it means absolutely nothing! To avoid confusion, the *level figure should always be qualified by a 'unit' term, e.g. voltage level,*



Gain vs. Level. An amplifier with 27 dB gain is fed an input signal whose level is -23 dBu to yield an output level of +4 dBu. The decibels of gain never require a suffix – it stands alone, e.g., 27 dB gain.

Approximations of True Loudness

The earliest "slow responding" meter was the VU meter, with a full wave rectifier, a reasonably-flat frequency response and 300 ms integration time. The best we can say is the VU is a tiny bit closer to true loudness than a peak meter. But it is very inaccurate. Rest in peace.

The rectified wave shown at right is a simplification of what needs to be much better specified in order to be an accurate measure of loudness. Even BS.1770 with its weighted mean square approach is a convenient, easy-to-calculate simplification of an actual psychoacoustic measurement of loudness.

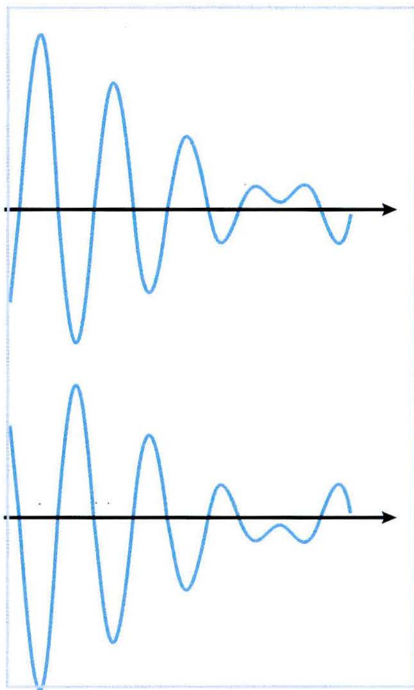
An actual measurement of loudness would be expressed in sones or phons. However, BS.1770 provides a convenient method of telling you how many dB up or down you may want to adjust your gain.

A discussion of proper filtering, time constants, and integration constants is far outside the scope of this book.

—Jim Johnston

sound pressure level, digital level. Examples: 40 dB SPL, -20 dBu, -25 dBFS. Each suffix defines the reference.

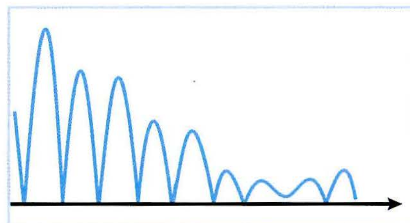
Average vs. Peak Level... Which of these two waves (pictured below) is louder? The answer is: Both have the same loudness. The first wave is identical to the second except its polarity is exactly reversed. I believe that the sound quality of transient peaks may be perceived differently depending on their polarity, but the perceived loudness of inverted material is identical to non-inverted. We consider the up-going direction as *positive* and the down-going as *negative*. The first wave has a maximum in the positive direction and the second has a maximum in the negative. However the *absolute value* of the greatest peak in each wave is the same, so each wave has the same maximum peak level.



Which of these two waves is louder?

Both waves also have the same average level, but the term *average* is a bit misleading because if we add up the positive and negative-going values over a long period of time, the *numeric average* must be zero, since in real life positive and negative waves decay to the static pressure of the atmosphere. Because the ear ignores polarity when it judges loudness, the loudness would be a little more correctly judged by the *rectified wave* (pictured below), where all the negative segments have been converted to a *positive*, or more correctly termed, *absolute* value. First we must rectify (get the absolute value) and then we can compute the average. (See Sidebar at left)

Loudness... is used specifically and precisely for *the listener's perception. Loudness* is much more difficult to represent in a metering system. In fact, it's best



Rectified wave

presented as a series of numbers rather than as one overall figure of "loudness." Exposure time and context also affect our perception: after a five minute rest, the music seems much louder, but then we get used to it again – another reason why it is wise to have an SPL meter around to keep us from damaging our ears. In an album, a loud downbeat that follows a soft song ending will seem much louder than one that follows a loud song ending. Since our ears are sensitive to contrast, our perception is variable. Keep this in mind when leveling an album. If we cheat the level up very slowly during a long, soft passage, we will often not notice how much louder the sound is getting since the ears have habituated. This is both good and bad: good if we can take advantage of it, bad if we take too much advantage of it.

Integrated Loudness (Program Loudness)... Integrated Loudness is officially defined by the ITU, and endorsed by the ATSC and EBU. It is often abbreviated **PL** — the *program loudness* of a program over time, expressed in *loudness units* relative to full scale. Despite human perceptual variability, the standards organizations have done the audio industry a great service by standardizing the term *program loudness* for broadcast and other purposes. An ITU/EBU program loudness measurement makes a calculation using sample level, weighting the frequency content, combining all channels, and integrating program duration. Two pieces of music that measure the same level on an old-fashioned flat level meter like the VU meter can have drastically different *loudness*. The standard ITU BS.1770-3 details how the calculation is made.

Program loudness is measured in **LUFS**, loudness units below full scale, the highest digital level that can be encoded, aka 0 dBFS. A stereo or surround program has only a single ("mono") LUFS measure; all the channels are combined according to their mean squared energy. A 1 kHz sinewave tone in *both* channels of a stereo program measuring -20 dBFS per channel will measure -20 LUFS on the loudness meter. A 1 kHz sinewave at -20 dBFS in only *one* channel will measure -23 LUFS. Mean squared combination ensures that interchannel phase relationships have no effect on the loudness reading. In fact, even if the two channels of a stereo recording are placed out of polarity and would normally cancel in mono, combining them according to their energy results in the same program loudness. LUFS is weighted according to K-weighting, but dBFS has no weighting (it is flat). When I use the term **hot CD** or **hot master**, I am referring to a recording which has a high measured program loudness and is probably highly compressed if it contains music with transient peak material. Our perception of the program's loudness is also affected by the behavior of the monitor DAC: if a program has peak distortion that would cause a certain DAC to overload, this DAC may appear louder due to the high frequency distortion. **Program loudness** would not have been a meaningful term in the analog era because analog tapes and LPs do not have a consistent reference. But with digital audio, 0 dBFS is always the same. Program loudness can also be measured in LU, which is loudness units relative to any arbitrary 0 LU, if full scale is less important than the 0 LU point itself.

Loudness Normalization... The process of correcting the level of a program to a standardized **target** level. It is simply a gain or attenuation value, a level control which can be applied on broadcast, or in the case of a media player like iTunes, just before hitting play on the selected song or movie. For example, the **EBU** has defined the standard target for program loudness for both European radio and television as -23 LUFS ± 1 dB. The **ATSC** has defined a -24 LUFS target ± 2 dB, so the two standards are reasonably compatible. This target is genre-neutral — it applies to all genres from spoken word to heavy metal!

This approach is not necessarily a fair esthetic way of judging how loudly a particular genre should be played; for example, we like to play electric music louder than acoustic music. But it is not possible to discriminate electric from acoustic by current artificial intelligence, nor has any previous media had this ability. In fact, the situation has been worse prior to the advent of loudness normalization. On Compact Discs, I can demonstrate folk music that sounds louder than heavy metal, string

MYTH:
"The red light came on while I was recording, but when I played it back, there weren't any overs, so I thought it was OK."
Contributed by Lynn Fuston

MYTH: *Peak Normalization Makes the Song Levels Correct.*

quartets that sound louder than symphony orchestras (due to peak normalization). At least, with loudness normalization, all genres will be at equal measured loudness, even if that is not the perfectionist's ideal.

At a minimum, normalization brings everything into line, and is not at all damaging to the program material—unlike the egregious processing still in place in U.S. domestic radio, some TV, satellite radio and some internet radio stations. The concept of loudness normalization actually encourages more dynamic mixing and mastering, because producers will no longer be motivated to try to make a "loud" album. As long as a suitable target level is chosen by the broadcaster or media player, any program will maintain its clarity and dynamics, yet still match in loudness to all other programs. Keep in mind that within any program or music album, the engineer can make some material as loud or soft as she wishes, as long as the average reaches the target. It might be a radio program combining music from different genres, sound effects and spoken word, and the program producer has the freedom to make the music louder than the speech, or produce heavy metal louder than Mozart! No matter what, there will be better sound quality than ever before in broadcasting.

When loudness-normalized media are in the majority, if you are creating a program destined for a targeted medium, such as television broadcast or iTunes Radio, it will pay to produce that program at the same level as the target, for it will be brought down during normalization if it is above the target. However, when some clients are demanding "loud" masters for the non-normalized media that remain, such as CD, then your program will likely be louder than iTunes' or broadcast target, and it will be brought down when it is played on normalized media. Be sure to hear how your material sounds normalized, and compare it to other material in the same genre that has also been normalized. For iTunes and iTunes Radio, this is easy to accomplish by turning on Sound Check in the preferences. For terrestrial and some Internet broadcasting, measure the PL. If it is above -23 LUFS, it will be brought down to that level. Lower level material that has been conservatively-processed can instantly sound better than squashed material, when broadcast on iTunes Radio or file playback with Sound Check turned on. Chapters 17-19 detail loudness normalization and how it changes our audio lives.

The great news is that BS.1770-3 is an international standard: it can be applied to all media, and (thank goodness) it has begun to permeate to all sorts of media, including audio for games, which has been standardized by the leading game manufacturers to a -23 LUFS target. iTunes' approximate target level is -16.5 LUFS, although they use another method to measure the loudness (and may change over to BS.1770 at some time in the future), so there is about a 2 dB variance from an EBU meter. The amount by which iTunes will adjust a given song can be found in the *get info window* within iTunes when **Sound Check** (Apple's normalization technology) has been turned on in the preferences. This is the value that will be applied during **singles normalization**, as opposed to **album normalization**, which preserves the relative levels of all songs in an album (a very good thing!). When album normalization is in effect, the loudest song determines the gain that is to be applied to the whole album. This is helpful in avoiding overloads, because it prevents the normalizer from raising the softer songs, which could cause peak overload. Keep in mind that if album normalization is in effect, pick the loudest song to find the amount by which the entire album will be adjusted.

iTunes' target is well-chosen and conservative: -16.5 LUFS accommodates the vast majority of recorded music with full headroom — the peaks are preserved without the need to add limiting or compression. I advocate that all music playback systems incorporate a target no higher than -16 LUFS, to follow Apple's conservative lead. Chapter 17 will cover iTunes Sound Check in more detail.

Integrated, Momentary, Short-Term Loudness... *Integrated loudness* is the same as program loudness (PL), which is the quantity you should be aiming for when taking a measurement of the whole program.

The EBU has defined two other time scales for use with loudness meters. The first is **Momentary Loudness**, abbreviated **M**, which is the loudness you hear *now*. M is averaged over a 400 ms period, which corresponds well with the VU meters many of us are used to. The second is **Short-term loudness**, abbreviated S, with a time window of 3 seconds. I wish they had come up with a different term, because it is not obvious that short-term is longer than momentary. It will be interesting to see which meter movement engineers gravitate to. It's also important to realize that the engineer might have to adjust the 0 value on the LU scale depending on which time window she likes to use, which I'll describe in Chapter 18.

Gating... The EBU standard R128 defines a measurement gate that purposely ignores soft program material more than 10 LU below the average, in order to prevent extra-soft passages and fadeouts from over-influencing the true average loudness measure. Most current EBU-compliant meters incorporate the gate. I recommend that all loudness meters for music measurement incorporate the gate.

Loudness Range... abbreviated *LRA*, is a well-defined statistical measure of dynamic range, essentially the difference between the highest and lowest gated loudness values in a particular program. It is the gating that permits LRA to become a valid and repeatable measurement. EBU tech document 3342 explains:

> LRA is defined as the difference between the estimates of the 10^{th} and the 95^{th} percentiles of the distribution of loudness. The lower percentile of *10%*, can, for example, prevent the fade-out of a music track from dominating loudness range. The upper percentile of *95% ensures that a single unusually loud sound, such as a gunshot in a movie, cannot by itself be responsible for a large loudness range.*

Additional gating is also performed. The statistically-minded can learn more about the LRA algorithm from EBU document tech 3342.

Esthetically speaking, the LRA value should only be considered to be a *very general guide*. A high LRA value has been misused by some broadcast authorities to reject perfectly valid program recordings, so the broadcasters still have much to learn from this new approach to loudness. As explained in EBU document tech 3343:

> Loudness variation is an artistic tool — it is the average, integrated loudness of the whole program that is normalized.

As a general example (*descriptive, not prescriptive*), LRA values found in recordings include: full movie: 10-25 LU, classical music: 8-23 LU, 80's pop/rock: 4-13 LU, TV show: 5-8 LU, Contemporary Pop music: 2-6 LU, Aggressive TV commercial: 1-3 LU.[3] Notice the wide variance. Some producers like a wide dynamic

PLR: A Convenient Fiction

Jim Johnston would like to point out that it is inappropriate to compare the *digital level peak measure* to *loudness* because the two quantities have different dimensions. Furthermore, the ratio of the single highest program peak to the average loudness is not microdynamics; microdynamics is the short-term changes in loudness. So a measurement of the amplitude of short-term changes would be the ideal. A continuous test signal with a high peak-to-average ratio reveals a weakness in the PLR method. However, programs that leave room for momentary peaks have a lower loudness and usually exhibit greater microdynamics and therefore a relationship between PLR and microdynamics. In experiments, Jim has found a reasonable correlation between a psychoacoustic measure of microdynamics and the simpler PLR method.

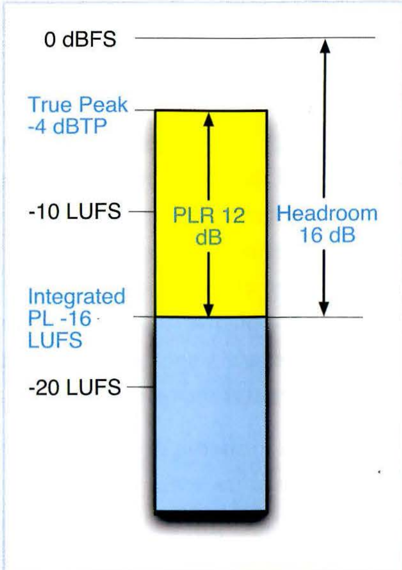
PLR is intended to be a guide, judged in a genre-specific manner by an experienced engineer. The Pleasurize Music Foundation uses a similar approach abbreviated DR. This has come to mean *Micro-Dynamic Range*. Now we're all on the same street, if not in the same house!

range, others like narrow. These are the early days of loudness normalization. At some time in the future we hope that the destination device (the radio, the computer, the portable music player) will be smart enough to raise low level passages in a noisy venue (such as the gym or your car), so that program producers can create material with a nice wide dynamic range that can translate via smart technologies to venues that require a narrow range. Bose AutoPilot® is already a very effective gain-riding technology that does not introduce obvious compression artifacts, and can raise the gain in a car when it's noisy, but prevent the sound from blasting at a stoplight, for instance. I hope this sort of technology migrates to portable media players, for that will mark the end of producers demanding that programs be made especially for drivers and joggers.

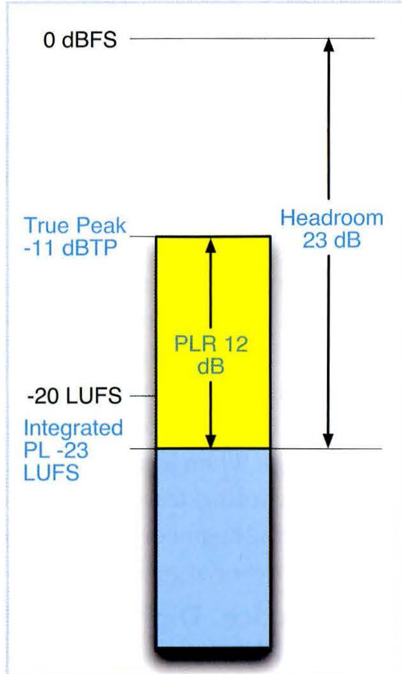
Sound Pressure Level (SPL)… is a measure of the amplitude or energy of the physical sound present in the atmosphere, expressed in dB relative to 0.00002 Pa (Pascals), which is defined to be 0 dB SPL. Originally this was determined to be the nominal threshold of hearing, although at certain frequencies we can hear below 0 dB SPL in a perfectly quiet room. For example, 40 dB SPL and 0.002 Pa represent the same pressure, the first expressed in decibels relative to 0 dB SPL, the second in absolute pressure units. 74 dB SPL is the typical level of spoken word 12 inches (30.5 cm) away, which increases to 94 dB SPL at one inch (2.5 cm) distance. While we often see language like *95 dB SPL loud*, this is both inaccurate and ill-defined, as *loud* refers to the user's perception, and *SPL* to the physical intensity. SPL measurements must include the weighting curve used, e.g., A, or C, the speed of the meter (slow or fast), and method of spatial averaging (how many mikes were used and how they were placed).

True Peak Level… another term newly-standardized by the ITU, is an estimate of the peak level that will be encountered at the output of a DAC or any other filtered process, such as an SRC. It's abbreviated *dBTP*. Compare this to **sample peak level**, which is the peak value of the digital sample, measured by traditional digital meters — but traditional digital peak meters are no longer recommended for program measurements. For example, the sample peak of a program might be 0 dBFS (the highest level any sample can reach), but the true peak could be anywhere from the same level to significantly higher, though 1 dBTP is about the most I've encountered with practical music. If the level is 1 dB over full scale, it is written as +1 dBTP. The + sign is optional. Although it is labeled "true peak," this is an estimated value. True peak meters incorporate a certain amount of oversampling so they will react similarly to a typical DAC, but since not every DAC uses the same amount of oversampling, certain variations are expected. That said, don't ignore the true peak measure: in fact, we must *embrace it* — for all practical purposes True Peak tells us that distortion will occur on playback if the level exceeds 0 dBTP. True peak is the modern-day over-level meter: it makes the sample peak meter and all previous attempts at measuring over-levels obsolete, whenever we are determining the level of a mix or master, and we don't want the consumer's playback DAC to overload, or downstream processes such as SRC to overload. In general, the more processing, distortion, or high frequency content, the higher the true peak will measure compared to the sample peak.

However, don't use true peak to assess the output of an ADC, because any over-level information will have been lost before the signal is captured. Rely instead on the manufacturer's efficacy in creating analog domain



Example music program: Program loudness in blue at -16 LUFS. True peak level at -4 dBTP. The difference between these two is the PLR, 12 dB. Headroom is defined as the difference between the program loudness and the peak capability of the medium, which in this case is 16 dB.



The same program loudness-normalized for EBU broadcast, with target of -23 LUFS. Note that the sound quality has been maintained, since the PLR remains the same. Headroom has increased. Since the 24-bit noise level is already inaudible, there is no perceptible signal-to-noise ratio difference .

over-detectors, or else be conservative when tracking, and don't exceed -1 dBFS on a sample-peak digital meter, preferably lower to be safe.

Headroom... The ITU standard permits us to define *headroom* in a more useable way than ever before: Headroom is the difference between the program loudness and the peak-level capability of the medium (0 dBFS).

Peak to Loudness Ratio... This new term, abbreviated **PLR**, is being introduced in the third edition of *Mastering Audio*. It has not been standardized by any authority. *PLR* is the ratio between the highest true peak not exceeding 0 dBTP, and the long-term average loudness of the song or album in LUFS. We must not use true peak readings above 0 dBTP when calculating PLR because higher readings would likely distort a DAC, so do not reflect the listener's perception. PLR is a reasonable indicator of a program's *microdynamics*, as described in Chapters 5-7. We could coin the term *high micro-dynamic range* to help describe a recording with excellent short-term movement. PLR is a rough numeric measure of sound quality, not a substitute for trained ears. It can assess whether a recording has sufficient transients or short-term dynamic movement. Until we have a direct measure of microdynamics, PLR makes a good substitute.

The above figures illustrate the relationship between program level, PLR, and headroom. When the program level is lowered during normalization, the PLR is maintained, and the headroom increases. *PLR* is a measurement of the program itself. In contrast, *Headroom* is the capability of the medium. In other words, PLR is an *actual* measure; headroom is a *potential* measure.

It is rare to encounter a piece of music with a PLR greater than 14 dB, and extremely rare to find one that reaches 20 dB, so this is the commonly cited maximum. A PLR of

"Embrace the True Peak Measure."

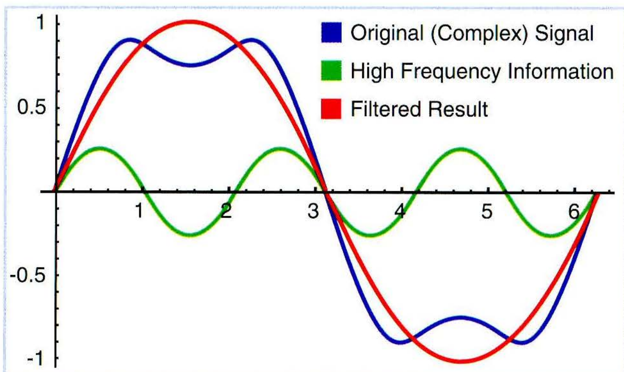
"Signals which cross domains can exceed 0 dBFS."

less than 10 (some might say 7 or 8) in a recording, especially one containing percussion, is likely to indicate overcompression, but ultimately the final judgment is by a trained ear, using a good monitoring environment. Even if we personally don't like the sound of results, the value judgment should be genre-related. However, be aware that normalization will expose abuses in PLR (caused by overcompression, clipping and peak limiting). When in doubt, please don't alter the PLR of a recording: leave these judgments to a trained mastering engineer, because when you lose sound quality, you cannot get it back.

II. Digital Level Practice, Digital Clipping, Peak Levels, Noise Floors

Why Does the Peak Level Go Up After Lossy Coding and Filtering?

The figure below shows that, contrary to what we might assume, *filtering or dips in an equalizer that we'd imagine would produce a* lower *output, can actually produce a higher output level than the source signal.* B.J. Buchalter explains:



Shown in blue is a complex wave. When the high frequency information (green) is filtered out, the result is a signal (red) that is higher in amplitude than the original! Image created by B.J. Buchalter.

> The third harmonic is out of phase with the fundamental at the peak values of the fundamental, so it serves to reduce the overall amplitude of the composite signal. By introducing the filter, you have removed this canceling effect between the two harmonics, and as a result the signal amplitude increases. Another reason for the phenomenon is that all filters resonate, and generally speaking, the sharper the filter, the greater the resonance.[4]

Jim Johnston notes that phase shift alone (common in equalizers, even without boosting) can cause an over on a signal which is close to peak. These phenomena occur in all filtered media and hardware: codecs, SRCs, DACs, when signals cross domains. This is why we have to watch out for overloads in succeeding stages or processes. Expect the peak level to go up!

Clip Measurements Are Deceiving

Ironically, some clipping can be inaudible, even after coding, but other clipping can sound quite obnoxious. I've mastered material with hundreds of measured clips which don't seem to have any audible effect. Jim Johnston explains why:

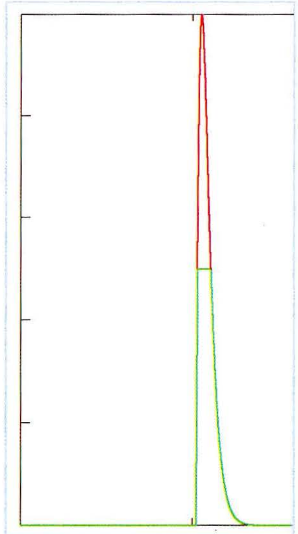
> Clipping can be divided into two categories: *pulse* and *tonal* clipping (illustrated on page 225). Figure A is a graph of the waveform of a short, positive-going pulse that has been clipped. The green section is the clipped portion and the red portion is the original signal, which has exceeded the clipping point.
>
> Let's examine the spectrum of the pulse's energy, which looks like Figure B. Notice that the spectrum of the distortion energy (in green) is

very close to that of the original pulse (in red), so it has been effectively masked by the original signal. The total sound is louder, but the added distortion is hard for the ear to detect. This is what we call "pulse clipping."

However, if we clip a tonal signal (Figure C), additional distortion components are created that are spread over frequency, are not masked, and are clearly audible and atonal to boot (Figure D) ...

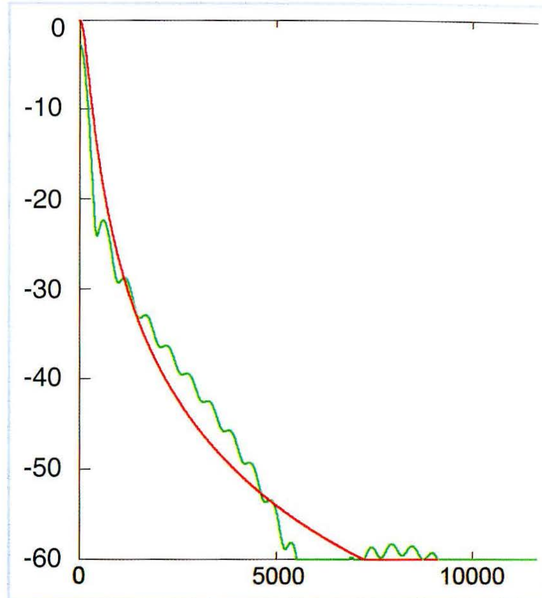
Figure C illustrates the waveform of the original 5 kHz sine wave in red, and the clipped sine wave in green. In Figure D, the spectrum of the clipped sine wave shows the original 5 kHz frequency in green, with all the added inharmonic distortion in red. Compare with the spectrum of Figure B, which does not show any egregious harmonics. In other words, *tonal clipping yields atonal results*. (I use "atonal" here to mean "having dissonant inharmonic distortion"). This demonstration shows that short-duration clipped percussive sounds are much less likely to sound bad than longer-duration tonal or musical

Figure A



Original short-term pulse signal (red). When clipped, it squares off (green). Clip demonstration images created by Jim Johnston.

Figure B



Spectrum of original pulse (red), spectrum of clipped pulse (green).

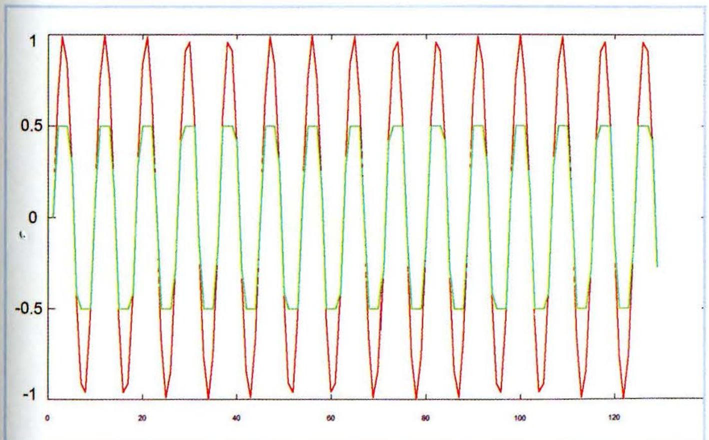


Figure C *5 kHz sine wave, unclipped (red), clipped (green).*

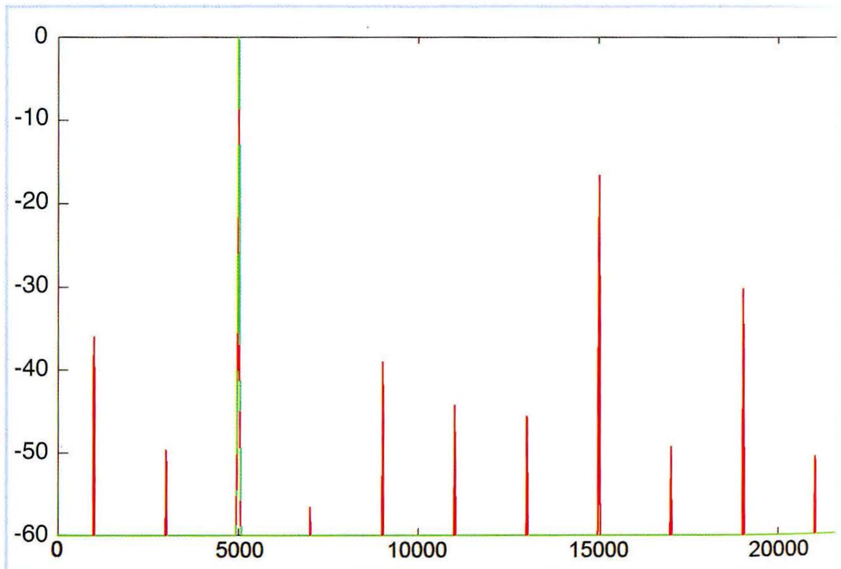


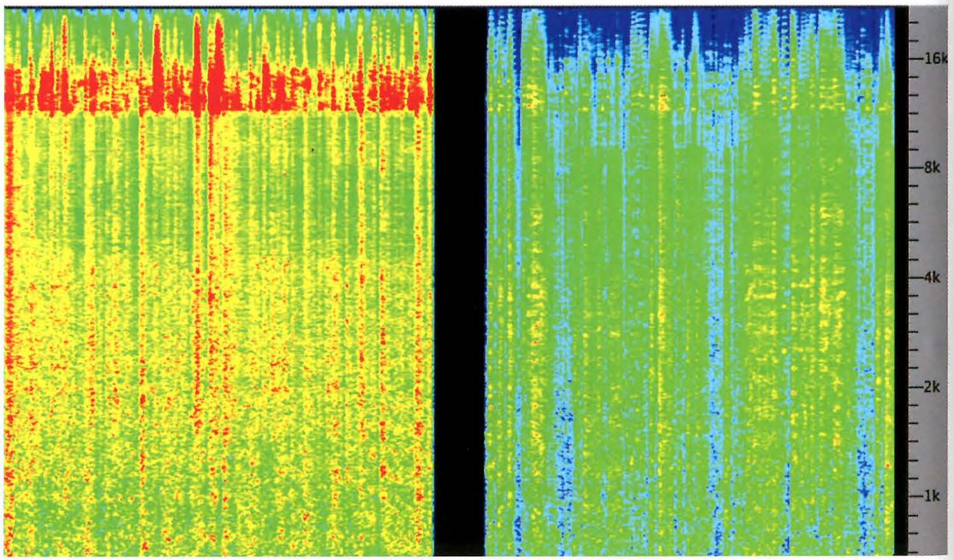
Figure D *Spectrum of original sine wave in green, clipped spectrum in red*

"Tonal clipping yields 'atonal' results."

sounds. When a music master is highly processed, equalized or compressed, it is more likely that if it has 0 dBFS+ levels, the clipping will be of the tonal variety, producing atonal results. In contrast, in conservatively processed masters, occasional short transient overloads, e.g., snare drum hits, may be invisible to the ear. We didn't start hearing about this problem, or at least the severity of it, until after the loudness race and the invention of digital processing, which can be egregiously abused. Even the oversampled brickwall limiter is not foolproof: I've discovered that very severe processing causes artifacts which such limiters are not able to ameliorate or prevent and can still make a consumer DAC overload unpleasantly. The best solution is to be conservative on levels, especially peak levels, but also RMS levels. For FM radio, or any other highly-processed medium, clipping of any type should be avoided (See Appendix II). I believe satellite radio is also processed and not simply normalized.

Clipped Material: Meet Lossy Codec!

After lossy coding, a PCM signal that sounds fine in the studio may tear up or sound quite bad, especially with the low bitrates encountered in digital satellite radio. I heard a recording on FM radio in which the artist's famous voice sounded like a parody of himself, as if he had a crunchy frog in his throat. I couldn't believe this was intentional, so I traced the problem. His CD had been mastered *stupid loud*. I measured true peak levels way above 0 dBFS, indicating that the material was already clipped. I decided to objectively compare what would happen when this CD was converted to mp3, versus when a non-clipped loud CD was converted. The result can be seen in the spectragram (pictured below) of the difference signal between the mp3 and the original CD. Time moves from left to right, levels vary from highest in red to lowest in blue, high frequencies are at the top. At left is the clipped CD master; at right is a "reasonably loud" master made without clipping, but with brickwall limiting. As you can see, the difference signal from the cleaner CD on the right is much lower in level: there are barely any yellow sections — most of the image is green or blue. The brickwall limiting has protected the right-hand master from overloading the mp3 conversion. But on the left, the red and yellow sections from the hot CD are the results of the clipping overloading the codec. All



Above left is a spectragram of the distortion created when converting a smashed and clipped CD to mp3. At right, the distortion when converting an "ordinary" loud CD which was brickwall limited and not clipped.

the harmonics that made the PCM recording seem loud in the control room have been converted to additional distortion after coding. This distortion is mixed in with the original signal. The horizontal red area circa 12 to 15 kHz represents severe codec aliasing distortion, very unpleasant to the ear. Be aware that dropping the level of the very hot CD before coding can only alleviate some of the damage, because there are still very high RMS levels that can easily fill up the Codec's bins and cause tearing distortion.[5]

Measuring for Clips, Certifying for MFIT

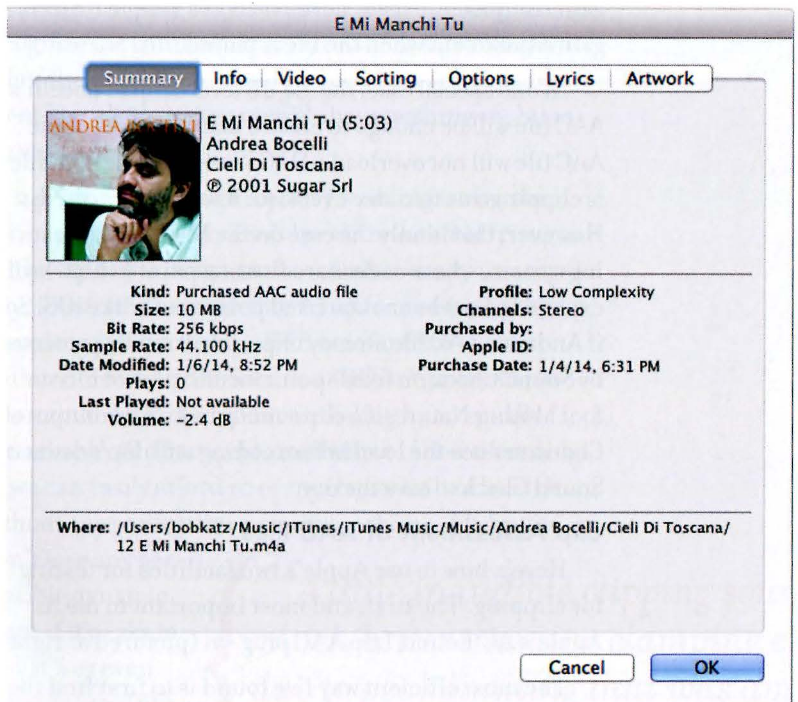
The Mastered for iTunes tools I described in Chapter 15 provide a way to measure a Codec's true peak levels. Sonnox offers a similar tool for Mac and PC: the *Sonnox Fraunhofer Pro Codec* plug-in. Even before coding, it's good while mastering to monitor true peak levels as well as loudness. We can master while checking a simulated codec, using either plug-in. But assuming we like the sound of the PCM master, the final stage is to inspect the peaks of coded material. Then we may decide to:

- lower levels until the codec does not produce over 0 dBFS or
- lower levels further if the codec sounds better, or
- at the engineer's discretion, keep the levels up and permit some degree of clipping in the codec, or
- insert additional peak limiting to attempt to reduce the clipping. But...

When listening, we may discover that, ironically, a little inaudible clipping sounds much better than the clamping effect due to peak limiting that was applied to remove the inaudible clipping!

This is a clear case of watching meters with our eyes and not listening with our ears. Fortunately, Apple is not the clip police, as Bob Ludwig points out. As long as an experienced mastering engineer approves the master, Apple will not reject it. But, regardless, there has been a lot of abuse and distortion in the loudness war, and it might have been better for Apple to have enforced an anti-clip rule. Ironically, clipping's savior will not be the mastering engineers, but rather Sound Check, because after normalization, peaks over 0 dBFS will be reduced and will not overload a DAC if the level has been attenuated and a floating-point codec used (I'll explain floating-point in a moment).

The whole issue of clipping will disappear with the end of the loudness race, because, as I mentioned, 16.5 dB is enough headroom above Apple's Sound Check target to allow the vast majority of existing music recordings to sound great without needing peak limiting, yet never hit full scale. Until our



The Get Info box in iTunes. Andrea Bocelli has entered the loudness race a little bit. E Mi Manchi Tu will be reduced by 2.4 dB when it is played on iTunes Radio. This is Loudness Normalization.

"Apple is not the clip police." – Bob Ludwig

clients get the message, we will have to watch and listen carefully for clipping.

If a Codec Clips, Will We Hear It?

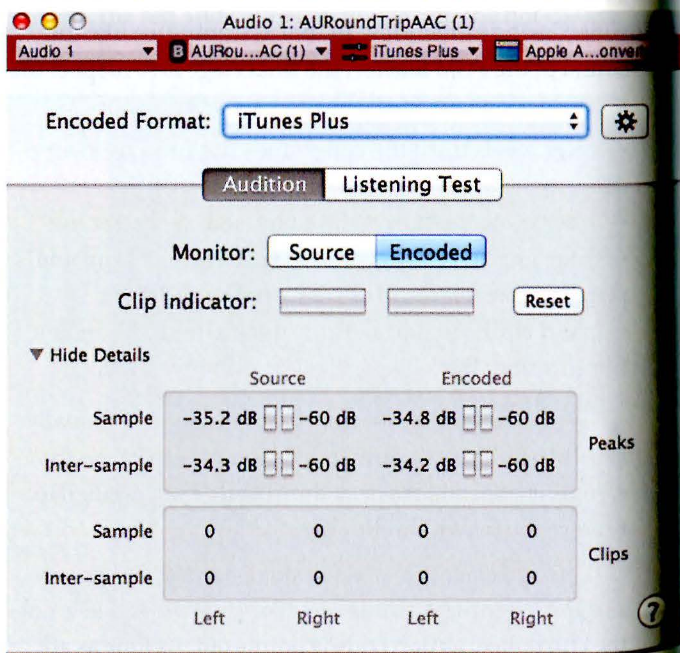
This iTunes "Get Info" box (pictured on page 227) shows that Andrea Bocelli's AAC file will be dropped by 2.4 dB when normalized, which is exactly what will happen when the tune is played on iTunes Radio. Remarkably, Internet radio, specifically iTunes Radio, is the first broadcast medium in which it is completely possible to predict how our recording will sound in the studio before it is broadcast. This is because, unlike FM radio stations, iTunes Radio does not add any compression, equalization, or processing to a recording. It is broadcast exactly as the file sounds, except that the gain is adjusted. This is performed as a simple (internal) gain adjustment when the file is played.

In one special case, the 2.4 dB level drop to Bocelli's AAC file will be enough to ensure that the peaks of the AAC file will not overload a DAC. So even if the AAC file is clipping due to codec overload, it will be protected. However, this is only the case on the Mac OS X operating system, where codecs are floating point (which I will explain below) but not on fixed point media like iOS. So if Andrea's AAC file already clips, it will not be protected by Sound Check on fixed-point media. It's not nice to fool Mother Nature: if a clip sounds bad on the output of a Codec, reduce the level before coding and don't count on Sound Check to save the day.

Clip Assessment of AAC files

Here's how to use Apple's two facilities for testing for clipping. The first, and most important to me, is Apple's AU Round Trip AAC plug-in (pictured at right).

The most efficient way I've found is to first find the highest peak in the song by the waveform, then listen to the encoded output in that vicinity. If it sounds clean and the loudest passage does not clip the AAC, I still try reducing the level (usually no more than a dB) to see if it sounds better. Often it does, for mysterious reasons having to do with the codec. If it does clip the AAC and sounds better with the clip removed, I calculate the attenuation necessary to remove the clip, then hope that the client will accept my judgment if they're concerned about competition (as they all are). Since codecs are a bit inconsistent, I drop the level another 0.2 dB to ensure that a subsequent playback doesn't clip. Next, I measure the file's Sound Check level by dragging the AAC file (extension is .m4a) into iTunes and checking with get info (command-I), just to see how competitive it will be on iTunes Radio. Or I can check even more information in the OSX terminal with the *afinfo* command. Less Sound Check attenuation is an indication



Apple's AU Round Trip AAC plug-in. It allows comparison between PCM source and AAC encode, and measures clips.

that the program does not have high RMS levels, that it is probably not overcompressed, and more likely the program will sound clearer than its competition on iTunes Radio. It helps to get the client involved. Ask them to turn on Sound Check in their iTunes, where they can hear the evidence for themselves!

Apple also likes to see the afclip report, even if we approve it despite the fact that there was clipping. Type *afclip* in the Terminal program, hit space, drag and drop the audio file into the terminal, then hit return. Here's part of an afclip report:

```
afclip : 2 ch,  44100 Hz, 'aac '
(0x00000000) 0 bits/channel, 0 bytes/pack-
et, 1024 frames/packet, 0 bytes/frame
 -- no samples clipped -
```

Or if there has been clipping, afclip reports the time and amount of each clip. Left channel is displayed with a hyphen. Here (at right) is an edited image of that report. There were about 20 clips reported in the vicinities of 259, 425 and 427 seconds, which are part of two issues that can be quickly auditioned and judged.

The Myth of the Magic Clip Removal

If the clipping has occurred in the PCM signal before AAC encoding, turning down the level while encoding will not remove the distortion, nor will normalization. Some mastering engineers deliberately clip the signal severely in an early stage, then drop the level slightly before AAC encode or making the CD master, so that the output of the master will not show any OVERs. This practice, known as **SHRED**, produces very fatiguing (and potentially boringly similar) recordings.[6]

Level Practice for Good Recording: The Truth about Signal-to-Noise Ratio

16-bit: The practice of hitting peaks to the top was introduced in the early days of digital recording, when

CLIPPED SAMPLES:

SECONDS	SAMPLE	CHAN	VALUE	DECIBELS
259.951417	11463857.50	2	-1.002338	0.020284
259.958486	11464169.25	2	1.000583	0.005059
425.276185	18754679.75	2	1.000956	0.008296
425.276190	18754680.00	2	1.000535	0.004643
425.276202	18754680.50	2	1.000526	0.004568
427.767001	18864524.75	2	1.001177	0.010221
427.767007	18864525.00	2	1.010359	0.089516
427.767012	18864525.25	2	1.011302	0.097619

```
total clipped samples
Left   on-sample: 0  inter-sample:  1
total clipped samples
Right  on-sample: 6  inter-sample: 24
```

OSX afclip report

converters were very non-linear at lowest levels. Those days are long gone, and now, even in 16-bit, we must recognize the truth: noise at -91 dBFS is significantly below the noise floor of rooms and mike preamps. We should be a little less paranoid about pushing 16-bit to the very last dB.

24-bit: Even though 24-bit recording is now the norm, some engineers can't break the habit of trying to hit the top of the peak meters, which is totally unnecessary, illustrated on page 230. As you can see, a 24-bit recording is remarkably 48 dB quieter than 16-bit (dither noise -139 dBFS vs. -91 dBFS). You would have to lower the level of a 24-bit recording by an equally remarkable 48 dB to yield an effective 16-bit recording! So we can easily afford 10 or much more dB of level drop without any perceptible noise from the 24-bit system. It won't lose any perceptible dynamic range if it peaks to -3 dBFS or even as low as -10 dBFS

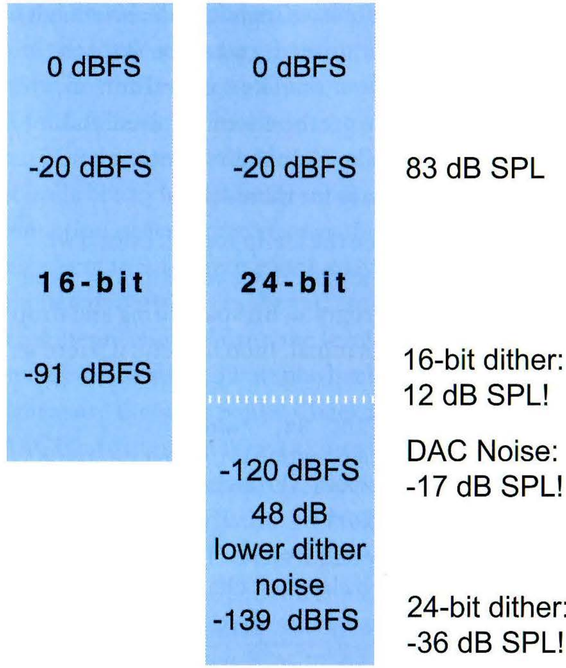
"A little inaudible clipping sounds much better than the clamping effect due to peak limiting that was applied to remove the inaudible clipping!"

MYTH OF THE MAGIC CLIP REMOVAL: *Turn it down after clipping and the clip will go away.*

—and you may end up with a cleaner recording, staying away from the distortion range of the ADC.

Let's imagine we're recording a violin today with forte passage at -20 LUFS (loudness) which "only" peaks to -10 dBFS. Tomorrow we'll be recording a snare drum at the same -20 LUFS forte which peaks to -3 dBFS. Should we record the violin at a higher level because there's peak room available? If we adjust our monitors to produce 83 dB SPL with our forte passage, the figures reveal that the dither noise of 16-bit would be at an extremely low 12 dB SPL, covered by the noise of all but the most quiet room. We'd have to crank up our monitor gain tremendously to hear 16-bit dither noise. Moving to 24-bit, 24-bit dither would be at an inaudible -36 dB SPL (that's "minus 36 dB"!), and the DAC noise at -17 dB SPL, well below the threshold of hearing. This demonstrates that it is absolutely unnecessary to raise the level of the violin—there is no SNR "improvement" to hear because there is already no system noise to hear! With the advent of the loudness meter, we can return to the classic way we recorded analog tape: Record all sources with 0 VU (or 0 LU) as forte and -20 LUFS, and the peaks will not overload because the medium has sufficient headroom.

This is the age of enlightenment: Install loudness meters on the console, align the system by sending a test tone at 0 LU, and adjust the ADC to produce -20 dBFS. This is the way we use analog consoles with digital recorders or DAWs: we send our console VU meters a test tone, adjust the gain of the digital recorder to equal -20 dBFS, and then we focus on the loudness meters (or the VUs if you prefer) and ignore the peak meters. We're protected with headroom and more than enough SNR, and we're going to make a great recording, just like the



A 24-bit recording would have to be lowered in level by 48 dB in order to reduce it to the SNR of 16-bit. The noise floors shown are with flat dither.

pros did back in the analog days. Keep repeating this mantra, "24-bit is 48-dB better than 16-bit" and you won't get into trouble.

If you hear any system noise at normal monitor gains when the DAW is stopped, you need to either fix an amplifier, or correct an error in gain structure. Mike preamps and room noise in a recording dominate its noise floor, not the digital medium. With dither and modern-day DACs, the distortion level is as low or lower than any analog system. This doesn't mean that I'm a digital audio lover per se, but simply that we shouldn't cast aspersions on digital audio's low-level performance, which is superb, as long as we follow the rules. Instead, we should look at the real issues of digital audio, investigated in Chapters 22-24.

Sales Gimmick?

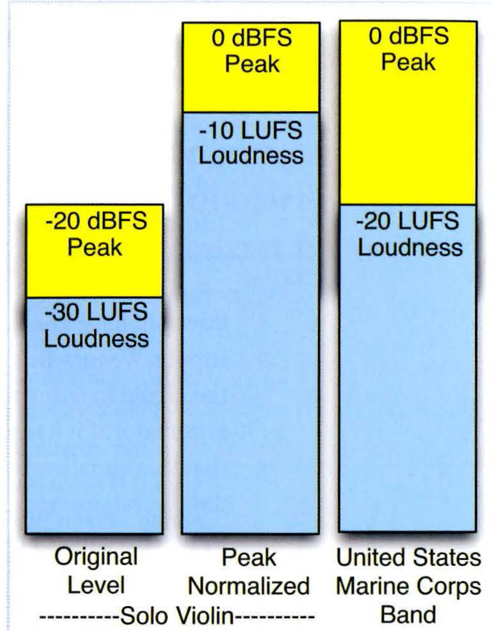
Some manufacturers advertise their ADCs as having available *additional headroom*, with a built-in compressor operating at the top of the scale. They claim that this gives a noise improvement, or a hotter recorded signal. In truth, this feature is only justifiable perhaps during an uncontrolled concert recording to protect from overloads — but otherwise, this is a sales gimmick. It's better to turn off the ADC's compressor and calibrate the record level with a loudness meter, as I have just described. If the warmth of the compressor attracts you, it makes more sense to add warmth later in mixing than to introduce mushiness during tracking due to an accidentally-overdriven compressor. It's hard to justify this practice.

Another sales gimmick comes from limiter developers who advertise that their limiters "improve resolution" by letting you raise the level when converting from 24 to 16-bit. Since the noise floor of 16-bit is already quite low, I can't rationalize adding additional peak limiting for the false impression of getting "resolution improvement," when limiting itself may cause a loss of transient clarity. Oh the irony!

The Truths About Peak Normalization

The Esthetic Truth: Since the ear responds to average levels, peak normalization completely distorts musical values. *Riddle:* How can you make a solo violin sound violently louder than the entire United States Marine Corps band playing forte?

Answer: Peak normalize the violin (simply raise the gain). Pictured (at right), a violin sounds naturally 10 (or more) dB lower than an entire military band, but after peak normalization, it's 10 dB louder! Now you know why we need to get rid of peak meters: They're deceptive, confusing, and even dangerous: peak normalization fueled the loudness war, as we will see in Chapter 17.



How to make a solo violin sound louder than a 100-piece military band? Answer: Just turn it up (peak normalize).

The Technical Truth: Mix engineers often mix all of their material to full-scale peak. Although this doesn't hurt, it offers no technical advantage, and the ballads sound louder than the rockers throughout the mixing process. Then, we mastering engineers just bring the level of the ballads right back down. The truth is that *material to be mastered does not need normalizing*, since the mastering engineer will be performing further processing. Clients often ask: "do you normalize?" I reply that I never use the computer's automatic method; I level by ear.

III. Gain Staging — Digital Chains

There is no loss or gain in a digital interconnection such as AES/EBU or S/PDIF, but we still have to be concerned about overloads. As we mentioned, equalizers can increase peak level even when dipping the level of a band! If an external processor overloads, try attenuating either the input or the output.

Headroom of the Processor. We can test digital systems for headroom, clipping, and noise using digitally-generated test tones and an FFT analyzer. Suppose we have a digital equalizer with several gain controls and equalization: we feed it a 1 kHz sine wave test tone at about -6 dBFS and turn up the 1 kHz equalization by 10 dB, observing that the output clips. Then we turn

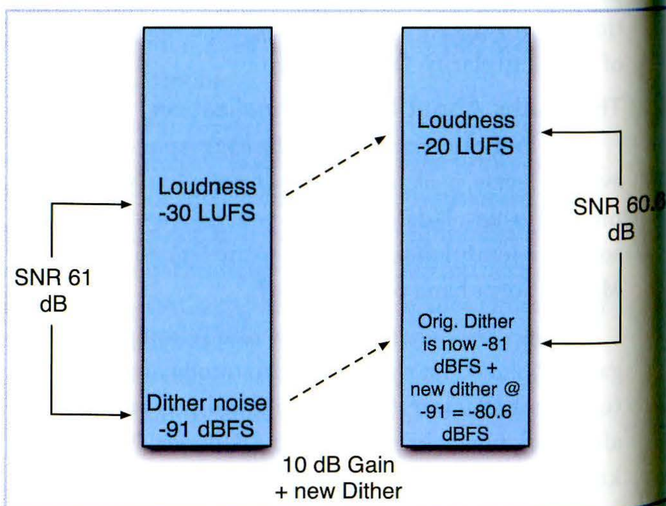
{"You would have to lower the peak level of a 24-bit recording by an astounding 48 dB to yield an effective 16-bit recording!"}

down the output gain control until the output is below 0 dBFS, and verify by listening or FFT measurements that the internal clipping stops. If the clipping does not go away, this tells us that the internal gain structure of the equalizer does not have enough headroom to handle wide-range inputs. We may be able to get away with turning down the signal in front of the equalizer, or the EQ's input attenuator if it has one, but the early clipping indicates that this equalizer is not state-of-the-art. Modern-day digital processors have enough internal headroom to sustain considerable boost in early stages, without needing an input attenuator: output clipping can be removed solely by turning down the output attenuator.

Noise of the 24-bit Digital Chain. With a digital processing chain, we no longer have to maximize the audio signal level in each piece of gear. Instead, we can send *pianissimo loudness* through a 24-bit digital signal chain without hurting the SNR, considering the inaudible (nominal -139 dBFS) noise of the chain. Processor internal resolution has generally improved so much that we don't have to be as concerned about cumulating digital processes. I now use a high-resolution processor in my monitor chain that has used thousands of calculations to create a complex filter—yet it still sounds pure and transparent. Regardless, I advise you to be diligent: test each digital processor and plug-in in your chain to confirm that it is capable of being bit-transparent and audibly transparent. Of course, analog simulators are intentionally not bit-transparent: they need to be tested with different criteria (See Chapter 22).

Noise Floors Add In Digital Audio Just As They Do in Analog

Noise floors sum in digital the same way they do in analog, including dither noise. Let's take an example of a 16-bit recording whose loudness is -30 LUFS (pictured below). In mastering, we may choose to raise its level by, say, 10 dB, and so must add another "dose" of 16-bit dither before turning it into a 16-bit master. Disregarding the mike preamp and room noise, the original 16-bit recording's dither noise is at -91 dBFS, and thus has a nominal signal-to-noise ratio of 61 dB (-30 - -91). We're taking a little liberty with this SNR calculation, since we cannot truly compare a weighted signal with an unweighted noise.[7] When we raise the signal by 10 dB, both the original signal and the original noise are raised equally, but the total noise floor rises. The new noise floor is the RMS sum of the original dither—which has become useless noise at -81 dBFS—and the added (new) dither, which is at -91 dBFS. The sum raises the noise floor to -80.6 dBFS, or 0.4 dB worse, which is an insignificant degradation. We are



Noise floors add in Digital Audio just as they do in Analog. The noise floor goes up 0.4 dB by sum of powers.

usually not so fortunate to be able to raise the gain of the source so much: the closer the gain gets to unity, then the power sum of the two dithers doubles by 3 dB—this may veil the sound quality a bit. It would be better to leave its gain at unity and not process the source. Luckily, noise-shaped dither can reduce cumulative sonic veiling; just don't push the noise-shaping far enough to change the tonality.

If we could avoid 16-bit dither by producing an output at 24-bit that the consumer could use, then mastering processing and gain-changing could be performed with no significant penalty, with a noise floor 48 dB below the noise of 16-bit. This is the promise of delivering higher wordlengths to the consumer, and another reason to record to 24-bit in the first place.

The World of Floating Point: A Primer

A **fixed-point processor** has a fixed maximum peak level of 0 dBFS and a fixed noise floor determined by its wordlength, which for dithered 24-bit is approximately -139 dBFS. But a **floating-point processor** is capable of doing tricks that do not relate to the real world. It is practically impossible to clip a floating-point processor: you can raise gain by hundreds of dB without clipping. A floating-point chain does not have a definable noise floor until it is converted to fixed point and dithered to the destination wordlength, but the lowest noise floor potential of a 32-bit float file is the same as 24-bit fixed—that is, -139 dBFS dithered. This internal resolution is retained, seemingly magically, regardless of the gain applied. In other words, you can attenuate a 32-bit floating-point signal by 100 or 200 dB, and it will still internally retain its 24-bit resolution. That's why we can drop the signal level by, say, 100 dB, store the signal as a floating-point file, open the file, raise the gain back 100 dB, and restore the original signal exactly! Most floating-point DAWs work in 32-bit floating point, which has a maximum coded wordlength of 24 bits, regardless of the gain (since its gain is determined by an 8-bit exponent and its data information by a a 24-bit fixed point mantissa). 64-bit floating-point audio resolution is becoming more popular in DAWs as CPUs become speedier, and so there is no performance or speed penalty.

Nearly all current native (CPU-based) plug-ins use floating-point processing. Probably 80% of current outboard digital processors use floating-point processing internally. However, *all ADCs and DACs use and require fixed-point data*, and that will never change, because at some point we have to deal with the physical world, which has a fixed noise floor.

An interesting wrinkle: The best current ADC chips have no better than -124 dBFS noise (A weighted), about 21 bits equivalent. However, some of the newest chips put out a 32-bit fixed-point word, which contains calculated data from their decimator accumulators. These are not "marketing bits" and should not be truncated.[10] It is the equipment manufacturer's responsibility to apply digital dither to take that down to 24 bits on the way out the interface. That is the correct way to deal with extra ADC data since most DAWs are not equipped to dither data on input. We expect the ADC to provide 24 bits because AES/EBU connections are 24-bit fixed point. So wherever floating-point data or fixed point data with greater than 24 bits meets "the real world," the signal must be regulated and dithered to 24 bits. A floating-point signal chain can only be maintained internally within a DAW—until developers take advantage of Firewire,

The 32-64 Confusion

When a developer states that their application or plug-in is "32-bit" or "64-bit" this means it is compatible with a 32-bit or 64-bit operating system, e.g., 64-bit Windows, or the latest versions of Mac OSX. A 32-bit application or plug-in can only address up to 4 GB of memory, while a 64-bit application or plug-in can address up to 16 EB (Exabytes) of memory! In practice, this means that you will never get an out-of-memory error when you run lots of plug-ins or apps, as long as you have enough physical memory. 8 to 16 GB of physical RAM is more than enough for a busy audio computer. This has nothing to do with the application's internal audio resolution. *32-bit or 64-bit internal audio resolution:* An application or plug-in can compute with 64-bit internal audio resolution no matter which operating system is used. This resolution has no effect on the memory space required by the app. It's a pity we're stuck with this confusion and must pay careful attention to what the developer tells us. Are we clear now?

Thunderbolt, or USB's capability of passing floating-point signal to external devices.

In a floating-point system, you can break all the rules I've taught you (sort of): literally ignore the individual levels in the chain and still it will not peak-overload. Most floating-point processors indicate when signal is above 0 dBFS, though internally they do not care. Some warn you with a red light that this signal level should not be fed to the real (fixed-point) world. You can test your DAW's internal signal chain for floating-point integrity by running the level of the first processor above 0 dBFS. Listen directly to that processor: it should sound distorted, since your DAC will overload. But if you then drop the level below 0 dBFS in the last processor of the chain, you should not hear distortion at the end of the chain.

This figure (page 235) illustrates that floating point really works: We connect the output of a Waves C1 compressor plug-in to a Waves L2 digital limiter. Notice that the output gain of the compressor (left side of figure) has been set to +6 dB. This has brought the highest peaks of the music into overload, which would clip a fixed-point system. The over indicators are in the red. But since the compressor works in floating point, it can show output levels greater than 0 dBFS on its meter, +3.5/+1.3 dBFS (left/right channel). Moreover, these values have been passed on correctly to the L2, as you can see under its threshold slider. With a threshold of -6 dBFS, an input signal of +3.5 dBFS causes a -9.5 dB gain reduction, as you can see in the L2's attenuation meter. The limiter keeps the output level to -0.3 dBFS or below by use of the output gain (ceiling), and no distortion will be heard (other than the dynamic artifacts of an extreme amount of peak limiting)! Please confirm that all your processors are interacting this well before "abusing" the signal chain. Otherwise, you might forget and feed a distorted signal to a real world Aux send — so it pays to watch for the red lights.

There is one further advantage to floating point over fixed. 0 dBFS+ signals produced by processes such as filtering can be reduced later in the floating-point chain without penalty. As I mentioned above, if a 32-bit float AAC Codec appeared to clip, and its output level is reduced in the floating-point domain before converting to fixed, the over-level can be completely fixed. Similarly, if performing sample rate conversion, save the result as a floating-point file. Then, even if the SRC level exceeds 0 dBFS, you can reduce the level of the floating-point result before converting to fixed, without penalty.

How can we tell if a processor or DAW is fixed point, except by calling the manufacturer? A fixed-point plug-in will audibly overload and its output levels will not exceed 0 dBFS on its meter. The only fixed-point DAW left standing (but already superseded) is Pro Tools version 9 HD, which cannot accept or import floating-point files. The only fixed-point plug-ins remaining in production are plug-ins designed to work specifically with Pro Tools 9 HD. Still, remember that the outside world and converters are all fixed point, so be diligent about inspecting all final output levels.

IV. Analog Studio Levels, Headroom and Cushion

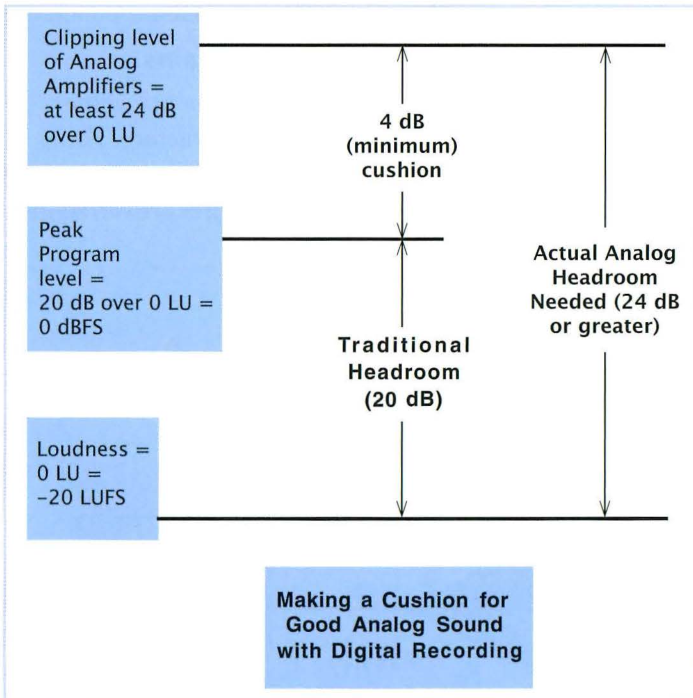
How to Protect the ADC from Clipping when Tracking: Some engineers still have analog tape practices on the brain, but we must break that thought pattern. Calibrate the ADC as I described on page 230.

Don't worry about "low" peak levels in this case, and don't let anyone tell you otherwise. A higher-level calibration is bound to get you into trouble when recording percussion. The only time you should be concerned about low levels when tracking is if you're recording a soft instrument that might have to be significantly raised in post production or mixing. Just raise the console gain to bring the instrument into the 0 LU range.

The Power of Floating-Point Processing: The first processor in the chain shows an "overload" (+3.5/+1.3). This level is cleanly passed via floating point to the Waves L2, which shows the identical "over" input levels. Yet it can still peak limit effectively and output a signal that is not clipped.

Headroom of the analog gear. Protecting your ADC and mix from clipping does no good if your analog console is distorting in front of the ADC! Not all analog gear is created equal, and the standard nominal +4 dBu[8] may be too high for two reasons:

The first reason is that a lot of cheaper analog gear described as "+4 dBu" may have a clipping point of +18 or +20 dBu, which is not enough headroom for the signal. This can be a big impediment to clean audio, especially since distortion accumulates when cascading amplifiers. The second reason is that it is good practice to keep peak levels below the clipping point, since some solid-state circuits begin distorting a few dB before they clip.[9] This means, *keep the music peak level below the analog distortion region, not just the clipping point*, to avoid the *solid-state edginess* that plagues a lot of solid state equipment. I suggest you adjust the nominal analog voltage level of your system (at 0 LU) to be 24 dB or



Provide a Cushion Above the Clipping Point for Good Analog Performance

MYTH:

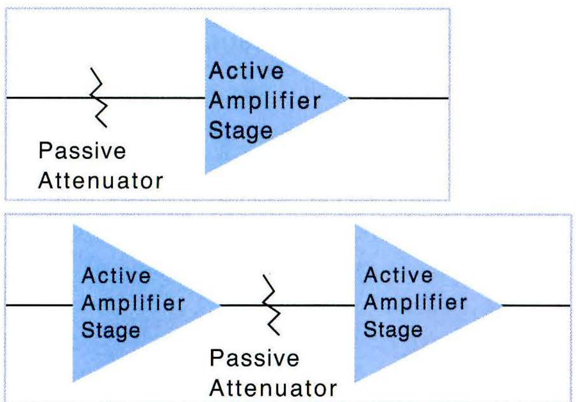
+4 dBu is always the best level to use for 0 LU with balanced analog electronics.

more below the analog clipping point of your processors or amplifiers. Measure the clipping point with an analog voltmeter and oscilloscope. You may find that your system should not be run at +4 dBu, but may have to be lowered to 0 dBu or below, in order to get a cushion above the clipping point (illustrated on page 235).

Internal clipping point in DAC. A similar issue is related to one of the most common mistakes made by manufacturers: to assume that, since the digital signal *clips* at 0 dBFS, it's OK to install a (cheap) analog output stage that would clip at a voltage equivalent to, say, 1 dB higher. This tends to produce a harsh-sounding converter or recorder, because of the lack of analog cushion and possible 0 dBFS+ analog levels. There is no easy way to test for this except to ask the manufacturer, or inspect the internal supply and components.

V. Gain Staging — Analog Chains

It's time to chain our analog equipment together, so we need to determine its internal structure. These figures (below) represent two possible internal structures. All complex equipment structures are variations on these themes.



In the top device, signal enters a passive attenuator and exits through an active amplifier stage. This circuit effectively has infinite input headroom. The bottom device's input headroom is determined by the headroom of the input amplifier.

To test analog devices, use a good clean monitor system, an oscilloscope, a digital voltmeter, and a sine wave generator that can deliver a clean +24 dBu or higher. The latter is tough to find, so you can create your own high-output oscillator by feeding it into a high-quality preamplifier and verifing that the output clips above +24 dBu. As shown in the figure, the first type of processor has a passive attenuator on its input, meaning that we can feed it any reasonable source signal without fear of input overload. We can tell if there is a passive attenuator on the input side by turning the generator up past +24 dBu and the processor's attenuator down, then seeing if the processor's output clips. If it can take at least +24 dBu without clipping, its internal structure really doesn't matter much because it must be a very clean design. We then test for noise by disconnecting the generator and listening to the output of the processor as we raise and lower the input attenuator. There should be no change in noise or hiss, and the output noise should remain well below -70 dBu unweighted, preferably below -90 dBu unweighted or A-Weighted. This is another indication that the device has a passive attenuator on its input. However, if the output noise changes significantly at intermediate positions of the attenuator, then the internal impedances of the circuit may not be optimal, or there may be some DC offset. The output noise of this device will be limited by the noise floor of its output amplifier. We determine the best *nominal operating level* of this device by taking the output clip point and subtracting 24 dB for headroom and cushion.

The second type of device has an active amplifier stage on its input, which is a more critical issue. It is very rare to find a solid-state device built this way that won't clip with >+24 dBu input. While you raise the signal generator, turn down the input attenuator to keep the output from overloading. If you hear clipping prior to the generator reaching +24 dBu, then the device cannot be used with a +4 dBu nominal signal. The clip point of the "weakest" processor (the one that clips first) determines the nominal analog level for your signal chain, at least 24 dB below this clip point. Also make sure that the output stage clips at a level no lower than the input stage.

System noise. When cascading analog gear, the noise of the system is determined by the weakest link. Set your monitor gain to be loud with a forte music level (for this purpose, 83-86 dB SPL), then stop the music and listen closely to the system noise floor at the last device in the chain. If the output of the chain sounds reasonably quiet, then the chain is decent. Unlike a digital signal chain, it pays to get the signal level in an analog chain as high as practical as early as possible in the chain, to keep the SNR high. In general, tube gear has a higher noise floor, so if gain has to be turned up, it should be in front of the tube gear, not after.

VI. Connecting the Analog and Digital Worlds Together

Standardize the Nominal Analog Level for Analog Gear

It's very important to *standardize nominal levels* in a studio; each piece of analog gear should have the same nominal voltage level, so when they are monitored or chained the result will be the same gain. All analog sources, consoles, tape machines, CD players, music servers, DVD players, and DACs should be adjusted to this level. ADCs and DACs should be aligned so that a sine wave at -20 dBFS produces this standard analog voltage.

Between the Devil and the Deep Blue Sea

Mechanical Meter Blues: One of the biggest problems in the contemporary audio studio is standardizing monitoring and mechanical VU metering levels, because the loudest production CDs are much louder than the masters that we hope to make, and would pin or damage any mechanical VU meter. Since we are using the same DAC for all playback, the best we can do is use a non-mechanical loudness meter with a variable calibration. In my studio, my loudness meters are calibrated most of the time for 0 LU at -14 LUFS to be somewhat competitive with a "reasonable" production product, and I've gotten used to seeing hotter meter readings when playing loud production CDs, AAC files or my own loud masters.

In Decibels We Trust

We've established a firm foundation — Now it's time to ring the liberty *bel!* Our trilogy titled *The Loudness Revolution* starts next.

1
Thanks to Jim Johnston (in correspondence) for helping to clarify some of these definitions.

2
IL (Intensity Level) = SPL - 0.16 dB. "For practical purposes, therefore, SPL and IL are numerically the same for progressive waves in air at STP." Blackenstock (2000) *Fundamentals of Physical Acoustics*, Wiley & Sons.

3
Thanks to Esben Skovenborg of TC Electronic for providing these statistics, presented in an EBU Technical Seminar 2011.

4
Thanks to B.J. Buchalter from MH Labs for this simple but powerful explanation and diagram.

5
The remnant distortion was graphed by subtracting the CD from the mp3 result. I had one client who was in the "make it stupid loud" department who spoke of "low class CD players" that distort, forcing us to turn down the level of his CD master. I tried to explain that excessive master levels were causing the problem, but he would have nothing of it.

6 Glenn Meadows and others discuss **shred**, on the Mastering Webboard:

Glenn: "Here's where I think all this is coming from, and it's kids-oriented. Ever pull up to a stop light, and get blasted from the car next to you? (I assume the answer is yes). Well, besides being aggravated, actually listen to what's going on. ALL of the audio is clipped and distorted on the high end. THAT's what people THINK things sound like, and are SUPPOSED to sound like.

So, for the artists and producers, who are used to "cranking it up in their cars," and having the top and transients clipped/distorted, if they DON'T hear that in their offices, then the mastering is just plain wrong. So, it's once again filtering back to the mix engineers, to provide that hash in the mix to satisfy their clients (remember, we ALL have to satisfy our clients first and foremost), so instead of losing the gig to someone else who WILL provide that edge, everyone is doing the same thing.

[Unknown respondent:] In other words, you are stating that the music business is currently conducted by people who don't know what a record should sound like.

Glenn: "You got it. Clean is OUT, distorted is in. If it's clean, it's not right. Unfortunately, I've had too many sessions go that way in the past few months."

Chris Johnson: "There's no future in that... clipping causes ear fatigue. Ear fatigue means listeners listen less before ceasing the listening. These people are only committing commercial suicide by going for stuff with no longterm sales capacity. It's just the same as if you put everything through an Aural Exciter turned up so far it really HURT, only this time around it's distortion."

7
Simplifying the arithmetic, we assume the loudness is at -30 LUFS, the dither noise is wideband at -91 dBFS (rounded from -96+4.77=-91.2). It is technically incorrect to subtract unweighted noise from weighted loudness to yield an SNR, although we must agree that the noise floor increases 0.4 dB by simple addition of the powers of the two wideband dither noises. The RMS sum calculation is:

Convert dB to power. $10^{\frac{-91}{10}} = 7.94 \times 10^{-10}$

Convert dB to power. $10^{\frac{-81}{10}} = 7.94 \times 10^{-9}$

sum of powers = 8.74×10^{-9}

10 log of sum = -80.58 dBFS

Thanks to Jens Jorgen for finding my error in the first edition and clarifying the formula for addition of powers.

8
The origin of using *+4 dBu* as a reference for analog audio instead of a more convenient number like 0 goes back to the earliest days of the telephone company. The reference used by the telephone company was based on *power*, with 0 dB at one milliwatt, which across their standard impedance of 600 ohms yields 0.775 volts. This reference is commonly abbreviated as *0 dBm*. The VU meter then came along; it is calibrated to produce a level of 0 VU with 0 dBm, but its impedance would load down the line and cause distortion, so the standard circuit added a 3600 ohm resistor in series with the VU meter, which attenuates the meter by 4 dB, so the circuit level has to be raised to *+4 dBm* to make the meter read 0 VU. In those days (1930), a headroom of 10 dB was considered necessary and sufficient, but fast forward to 1980+ and 20 or 24 dB headroom requires more extreme power supplies and circuits than we really would need if only the U.S. would lower the absolute level of the systems, like the more-sensible Europeans. Regardless, the *dBm* has evolved into the *dBu*, which means *decibels unterminated*.
Nowadays, modern-day equipment generally has low impedance outputs and high impedance inputs, so the old power reference has no meaning. But to keep using the same ancient Telephony-based levels, we kept the historical reference of 0.775 volts instead of a more convenient number like 1 volt! Now when the dB is referred to a voltage of 0.775 volts, we call that **0 dBu**. And to make a VU meter read 0 in a modern low impedance circuit with the right resistors, we have to feed it +4 dBu, or 1.23 volts.
The equations are:
If 0 dBu is 0.775 volts, then +4 dBu is 1.23 volts. 20 * log (1.23/.775) = 4.
I thank Mike Collins for reminding me to include this explanation.

9
This is also dependent on the skill of the designer. Some IC operational amplifiers perform very well up to the clipping point. Power supply design and regulation has a lot to say about sound quality near the clipping point. To avoid the nasties, use conservative levels, measure and listen.

10
Thanks to Bruno Putzeys for pointing that out (in correspondence).


"We'll FIX it
in the
Mastering."
—Anon

CHAPTER 17

The Loudness Revolution: The War is Ending

I. The Loudness War

History

There has been a loudness war (also known as a *loudness race*) since the dawn of commercial recording at the end of the 19th century. At first a producer's motivation to make louder records was to overcome the poor signal-to-noise ratio of Edison cylinders. In the acoustic era they could make things louder only by getting the performers to play as loudly as possible and by using acoustic amplification, which required big recording horns.[1] Then, from 1925, the electrical recording age began using microphones and amplifiers instead of acoustic horns, raising the bar and instantly obsoleting all the previous acoustic records — not just because the new records had a better signal-to-noise ratio, but because they sounded louder and clearer with better high frequency response. When the long-playing vinyl record debuted in 1948, the noise of the recording medium had been sufficiently reduced so that engineers could achieve an impressive amount of dynamic range and impact. In 1982, the introduction of the Compact Disc created the promise of a medium with an inaudible noise floor and potentially greater dynamic range. But little did we know: The CD presaged a highly-accelerated new loudness war, and by the year 1995, we had come full circle. We began to make popular music recordings with less dynamic range than a 1909 Edison Cylinder! How this came about is an important story to tell in this chapter. Even more important, we need to learn how audio engineers are changing digital systems so that a loudness war like this can never happen again.

Four different music recording styles (see text).

Our story is not just about increasing loudness, but about dynamic range loss, distortion, ear fatigue and possible loss of sales as customers get tired of fatiguing-sounding music. Our primary goal as audio engineers is to serve the music. If that isn't happening, we need to find new ways to restore the balance — and, as we'll see, loudness normalization is a critical part of that quest. Keep in mind that it is the artist's and producer's prerogative to produce any sound quality they desire. However, this Chapter will make it clear that the loudness war pressures artists to produce distorted recordings, even if that was not their desire. It doesn't have to be that way.

Here are four waveforms from a digital audio workstation, showing four different styles of music recording (pictured above). The more dense the waveform, the less the music's dynamic range and PLR. See Chapter 16 for definitions of these new abbreviations. Track 1 is a piece of heavily-compressed pseudo "elevator music" I constructed for a demonstration at the 107th AES Convention. Track #2 is John Mellencamp's *Love and Happiness* (1991), with a program loudness of -12.5 LUFS, true peak of +0.2 dB, an LRA of 5.2 dB, and a remarkable PLR of 12.5 dB, an indication of its punch, life, and impact. Track 3 is Ricky Martin's *Livin' La Vida Loca* (1999), whose PLR is about half of the 1991 CD, only 6.7 dB; its program loudness is -6.7 LUFS, about 6 dB louder than the 1991 CD. Its true peak is +1.2 dBTP, distorting any DAC; its loudness range is only 3.4 dB, which is truly *una vida loca*. In contrast, track 4 looks very dynamic, with an LRA of 6.2 dB, 3 dB more dynamic range than *Livin' Loca*. Amazingly, track 4 was recorded 87 years earlier, in 1912! It is an Edison cylinder of a military band playing southern plantation songs. Has sound recording technology advanced more than 100 years just so we can make elevator music and turn it into square waves?

The Origin of Loudness Envy

The origin of "loudness envy" is simply psychoacoustic: when two *identical* programs are presented at slightly differing loudness, the louder of the two appears to sound "better," and therefore attracts listener attention. When a producer's recording is more than 2 dB lower than another, it typically causes him to have second thoughts about the level. This makes some sense, at least until the sound quality changes the artist's intent. It is ironic that many established artists' recordings, including Michael Jackson's, have gone through several remasterings over the years to raise their level to the contemporary standard for each period. For example, Michael's song *Beat It*, from the *Thriller* album, was mastered to CD four times, first in 1982 with the original album, when it had a program loudness of -18.28 LUFS. Three more remasterings followed, raising the loudness with each re-release, until by 2009, the track had a PL of -7.35 LUFS (on the compilation album *This is It*) — a gain increase of more than 10 dB (twice as loud, and much more compressed).

Figure A

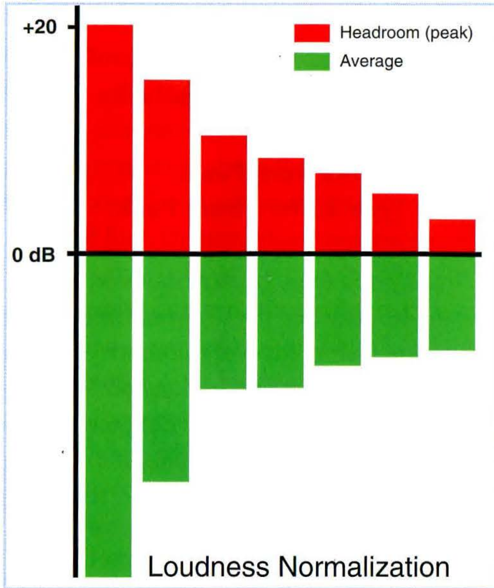


Figure B

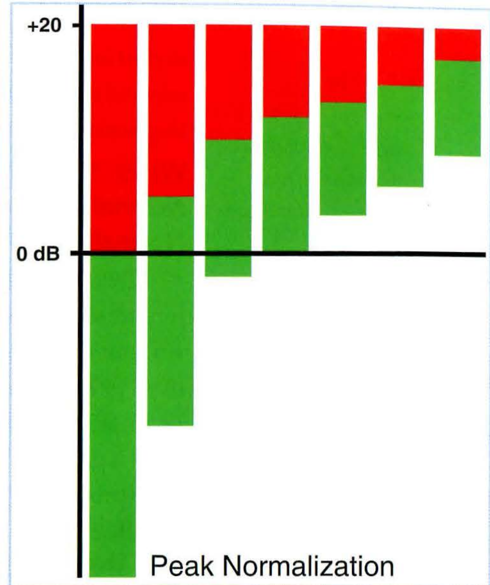


Figure A: Analog recordings were once made similar to this. A standardized forte level of 0 dB yielded fairly consistent loudness from album to album, regardless of the peak level.

Figure B: ***Pandora's Box: The Fuel For the Loudness Race.*** *Digital technology lets us normalize to the peak, giving an extreme artificial loudness advantage to highly compressed material.*

20 dB Louder in 30+ Years!

But, as they say, you ain't seen nothing yet. In 2014 a commercial release was discovered with an average program loudness of +1.9 LUFS, previously thought impossible to achieve, 20 dB louder than Michael Jackson's *Thriller — effectively the difference between a shout and a whisper*. In this situation the consumer puts on an old CD, followed by a newer one, and blasts his ears as well as his loudspeakers and neighbors. If you think *Loca* sounds distorted, just listen to a recording with a program loudness hotter than digital full scale (better you shouldn't)!

Peak Normalization vs. Loudness Normalization

Why has the digital race been so much more intense — and so much more damaging — than the analog race? How did we arrive at a 20 dB difference between the loudest and the softest pop recording? Even in the heyday of the LP loudness race, the difference between the loudest and softest pop LP was no more than about 6 dB, which is perfectly manageable. There must be a technical explanation.

It can be said that the engineers who produced LPs were trained professionals who did not abuse their expensive equipment. In the analog recording system, most engineers used average-reading (VU) meters, ending up with a fairly consistent average level and loudness regardless of the amount of compression used. The LP would either skip or cause extreme sibilance if the average or the peak level was made too high, so its loudness was self-limiting, or worse, the expensive cutterhead would burn up. Therefore, the way LPs used to be made is similar to loudness normalization (Figure A, above). The first bar represents a recording with a high (20 dB) PLR; its loudness in forte passages is in green at the 0 dB line and its peak level is in red at the top of the bar. Reading from left to right, we see that the forte passages of each album are placed at 0 dB, leading to a fairly consistent loudness from album to album, regardless of the amount of compression. The right-hand bar represents a recording with extreme compression and decreased PLR (less red area). However, note that the more compressed the recording, the greater the distance between its peak level and full scale. What would happen if we raise that peak until it hits full scale?

This is exactly what engineers began to do. Digital recording allowed us to peak-normalize, which opened up *Pandora's Box* (Figure B). If we compress and peak-normalize, the recording's loudness goes up as we decrease dynamic range, seen in the bars from left to right. Since digital mastering engineers can easily normalize to the peak, highly compressed material gains an extreme level advantage over uncompressed material. *The Compact Disc became the catalyst for the accelerated*

Vicious Circle

1. Engineers peak-normalize. The acoustic advantage makes peak-normalized acoustic material sound too loud compared to electric material.
2. This makes producers of electric music feel inadequate, so they compress and raise their level.
3. This challenges the acoustic producers, firing the next salvo — overcompressed acoustic music.
4. Around the circle we go again...

digital loudness race by allowing engineers to apply unprecedented levels of peak-normalization — the fuel that keeps the motors running.

Which Type of Music Benefits from Peak Normalization? Which Type Suffers from Peak Normalization?

The structure of sound gives some types of music a loudness advantage over others when peak-normalized, without applying any compression. Most of the short-term peaks of music are transients of percussion and percussive instruments. As I demonstrated in Chapter 16, it's possible to make a solo violin sound louder than a 110 piece military band, simply by raising its level! It's possible to make Joan Baez or Bob Dylan sound louder than Metallica, a harpsichord sound louder than a grand piano, and a string quartet sound louder than any combination of strings with percussion. Acoustic works that don't have much percussion don't suffer when they are raised, because their PLR is low enough to allow them to be raised without adding compression (music with a short red bar in the figures). I call this phenomenon the *acoustic advantage*. Another aspect of the acoustic advantage is that listeners prefer to reproduce music at its natural sound pressure level.[2] A soft singer sounds just fine played soft next to a rock band played loudly, and she sounds too loud when her loudness is made equal to the rock band. It's not necessary for an acoustic recording to be at equal measured loudness in order to compete with an electric one.[3] This is why most acoustic recordings don't need to be peak-normalized in the first place! The loudest master I ever made was of an uncompressed, close-miked solo pennywhistle, which the client insisted on peaking to full scale. This pennywhistle recording sounds louder than a hyper-compressed and distorted rock CD. So the beginning of the digital loudness war was an innocent move by engineers reading their peak meters: they placed the peak levels of acoustic music at full scale. This happened very early, by about 1984, and turned out to be the first mistake.

This was the beginning of a vicious circle. The acoustic advantage made the rock producers very mad, as their existing recordings then sounded much too low, by as much as 6 to 8 dB! Two examples I like to cite are the earliest rock releases on CD, Michael Jackson's *Thriller* and Black Sabbath's *War Pigs*, great-sounding recordings that are at least 10 dB lower than later rock recordings. This is despite the fact that both recordings originated on analog tape, which already compresses peaks by 6 dB. You have to turn up your monitor control to play these early releases at their intended loudness (and there's nothing technically wrong with that).

The natural next step was for the rockers to apply mastering compression, which let them raise the program loudness, but also softened the attack of the drums and made the image sound smaller. As long as this compression made the sound beefier, fatter, and punchier, there was no reason to complain. However, if a rocker liked a more open, clear sound, and discovered his album was 2 dB softer than the competition, he may have decided to compromise his standards in the name of competition, and he lost his signature sound. For example, Lyle Lovett's *Joshua Judges Ruth* album is much more microdynamic and open-sounding than his later works, possibly due to the loudness race.

A style that initially benefited from peak normalization was hip hop, which even in the late 80s employed synthesized bass drums that do not exhibit large peak excursions (they are naturally compressed). Hip hop style contributes to the loudness race because low-PLR sampled instruments instantly sound loud when peak-normalized. However, it didn't take long for the next salvo to create louder hip-hop by simply adding some production compression on top of these samples, raising the bar and producing a grittier sound that required all hip hop producers to follow suit or they would be more than 2 dB below the gritty-sounding leaders. For hip hoppers, loudness is of prime importance: a 2 dB loss is considered serious.

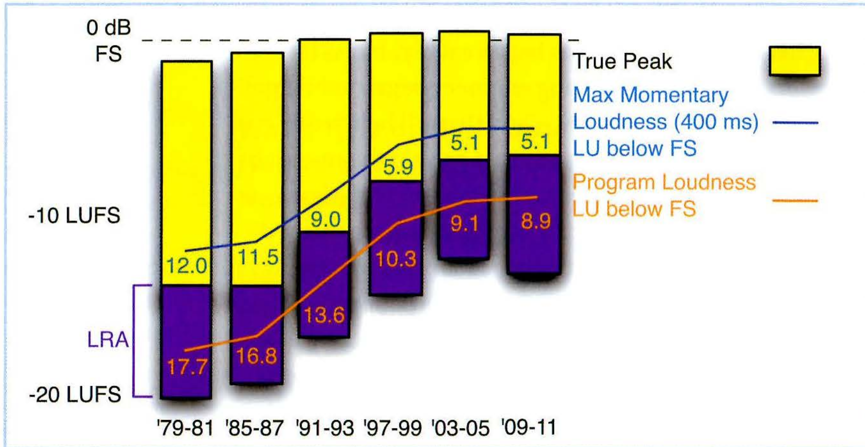
A rock mix with a snare drum mixed loudly inherently sounds louder, so mixes with a bit more aggressive snare immediately became more competitive. When this loud snare mix was combined with aggressive compression, it stepped up the bar so much that every other style (especially hip hop) had to do the same loud snare trick in order to compete, painting the other mix styles into a corner. One extreme example of this is *Green Day*, whose *American Idiot* album is very hard to compete with. It attracts a lot of instant attention. It's hard to equal in loudness, unless the competition also has an excellent recording, the same style mix, the same style music, and the same style of mastering. No wonder there are so many imitators but very few are successful!

Some genres suffer greatly from trying to make them sound loud. Right from the first year of the CD, genres such as Latin-Jazz, Salsa, and other styles that depend on hot, snappy percussion, would (and did) suffer when producers tried to make them sound as loud as gringo rock: they lost their snap, punch and space. It took a lot of effort to convince the A&R director of a latin record company that the masters I was making of classic salsa hits should not be made too loud. Loudness is a drug!

Documenting The Loudness War

The digital loudness war began calmly, but as it accelerated, every mastering engineer began to bring program peaks to full scale — and then all hell broke loose. For this part of the account I rely on the research of Rudi Ortner, whose measurement and statistical analysis of over 10,000 charting (best-selling) recordings documents the loudness race more accurately than anyone has done before.[4] Ortner's paper covers the years between 1951 and 2011, but first we will concentrate on the years of the Compact Disc, since about 1982, and for graphic clarity we'll skip every other three-year period. Ortner's statistical analysis is the *median* of all the data measured during each period. Median is a statistical way of computing average, it is the middle value of a distribution of values: half the values lie above it, and half below it.

In the figure on page 246 we graph the program loudness of all top selling songs for each 3-year period (orange curve). Program loudness (which corresponds approximately with mezzo forte) increased from -17.7 LUFS to -8.9 LUFS, a loudness increase of 8.8 dB, almost twice as loud, in 30 years! The maximum momentary loudness (MML), which corresponds approximately with double forte (blue curve), increased from -12.0 LUFS until, as we can see, it leveled off to -5.1 LUFS, not because engineers didn't want to go louder, but because the processors they were using had reached saturation. The ratio between program loudness and MML therefore decreased from 5.7 dB in the first period until, 30 years later, it was only 3.8 dB — betraying the use of greater downward compression. Forte no longer sounds like forte when it's not much louder than mezzo forte. This progression continues past



The increase of loudness of charting CDs throughout 30 years. *Program loudness (orange curve). Max momentary loudness (blue curve). Loudness range (violet bar). True peak level (top of yellow bar). Based on Ortner (2012. All plots are median of data for each period.)*

Ortner's survey: Three years later, the record-breaking 2014 CD's MML was an astounding +3.5 LUFS (I didn't think anything could get that hot)!

Ortner's data shows highest true peak levels (top of yellow bar) through 1996 didn't hit higher than +0.1 dBTP full scale, but by 1997 the *highest true peaks* exceeded full scale by about +0.5 dBTP. Anything above 0.1 dBTP is quite distorted. Keep in mind that distortion accumulates in every stage: initial lossy coding (AAC, mp3), cumulative lossy coding (when an AAC file is subsequently broadcast with its own codec), DAC overload, consumer DSP processors, and so on. Ortner measured many notably-high individual true peaks. These are only three examples: 1994, +3 (Bon Jovi's *Prayer*); 1999, +4.44 (Nine Inch Nails *Where is Everybody*); 2005, +3.86 (Madonna *Let It Will Be*). The result: harsh digital distortion, loss of headroom and squeezing of dynamics. The true peak of the extreme CD (2014) is +5.4 dBFS! Let's hope things come down from here on.

However, the loudness range (Figure A, page 247) remained fairly constant throughout the 30-year period. The figure shows that LRA remained fairly constant throughout 30 years, and even increased slightly in the last period, leading some observers to conclude (wrongly) that there was no loudness race.

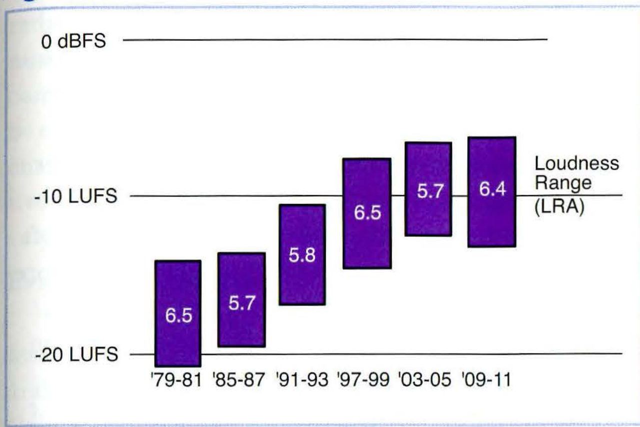
The real smoking gun is the insane reduction in PLR. PLR is a measure of compression/limiting. A small PLR means the recording likely has less impact and punch (Figure B, page 247). We see a big decrease in median PLR throughout this period by 7.7 dB: from 16.6 dB in the first period until 8.9 dB 30 years later. Note that we limit the true peak used for calculating PLR to a maximum of 0 dBFS, since DACs would be into distortion above that level.

Dynamics Processing Required to Achieve a Given Program Loudness

Ortner measured the integrated program loudness for each song, then he computed the median of all the PLs for all the songs in that period. For example, in the period 2009-2011, the median program loudness is -8.9 LUFS. But some of the highest values of integrated program loudness that Ortner found include: -4.82 (Reptile *Them Crooked Vultures*), -5.37 (Green Day *Horseshoes and Hand grenades*). It is impossible to produce those kinds of levels without using extreme processing and inducing harsh digital distortion.

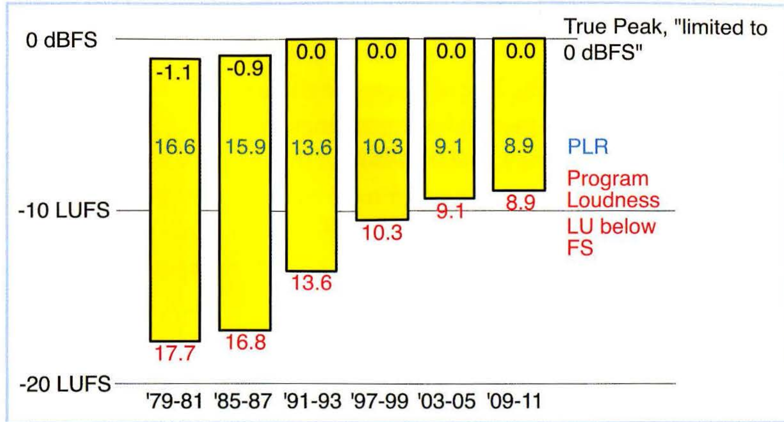
Figure C is an estimate of the kind of processing that is required to achieve the levels at each stage of the loudness war. It demonstrates how the loudness war pressured artists to distort sound and use extreme compression, even if that was not their artistic choice. As we can see, processing began gently in the first period and eventually reached a point where the loudness ceased to rise at the same rate, but the distortion increased dramatically. At the left side of the chart, circa 1980, little

Figure A



Median Loudness Range *remained fairly constant throughout 30 years. The violet bars show that loudness range remained about the same throughout the 30 year period, and even increased slightly in the last two years (based on Ortner).*

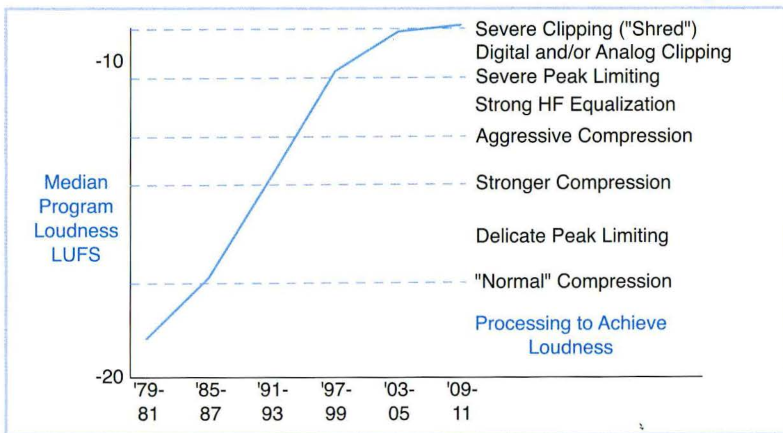
Figure B



The Insane Decrease in PLR *throughout 30 years between 1979 and 2011 (based on Ortner).*

or no mastering processing means little or no added distortion. Then, to raise a program's level, we began to use some compression. In the next period, peak-limiting entered the picture (it would rarely have been necessary if not for the loudness war). Next we added stronger and stronger compression until the sound reached saturation and lost its punch. One impetus for this was caused by the mix engineers, who in the 90s started to imitate the level and sound of already-mastered music. That inspired the mastering engineers to push even harder so our masters wouldn't appear lower than the mixes! So we resorted to strong high frequency equalization. This makes things seem louder because of the equal-loudness contours, as well as bright and fatiguing. Eventually we employed both digital and analog clipping. At the top of the figure, egregious processing represents mastering engineers' attempts to beat the digital system by raising the peak level over 0 dBFS, creating severe peak overloads and distortion that overloads encoders and radio transmitters. Notice how the slope of the curve of loudness increase flattens, showing that aggressive processing brings diminishing returns in intrinsic loudness, to say nothing of the sound degradation. Bear in mind that the whole idea of a "loud master" is only a conceit, since the consumer is in charge of their volume control and, more importantly: *"extra loud" masters sound wimpy when listeners adjust*

Figure C

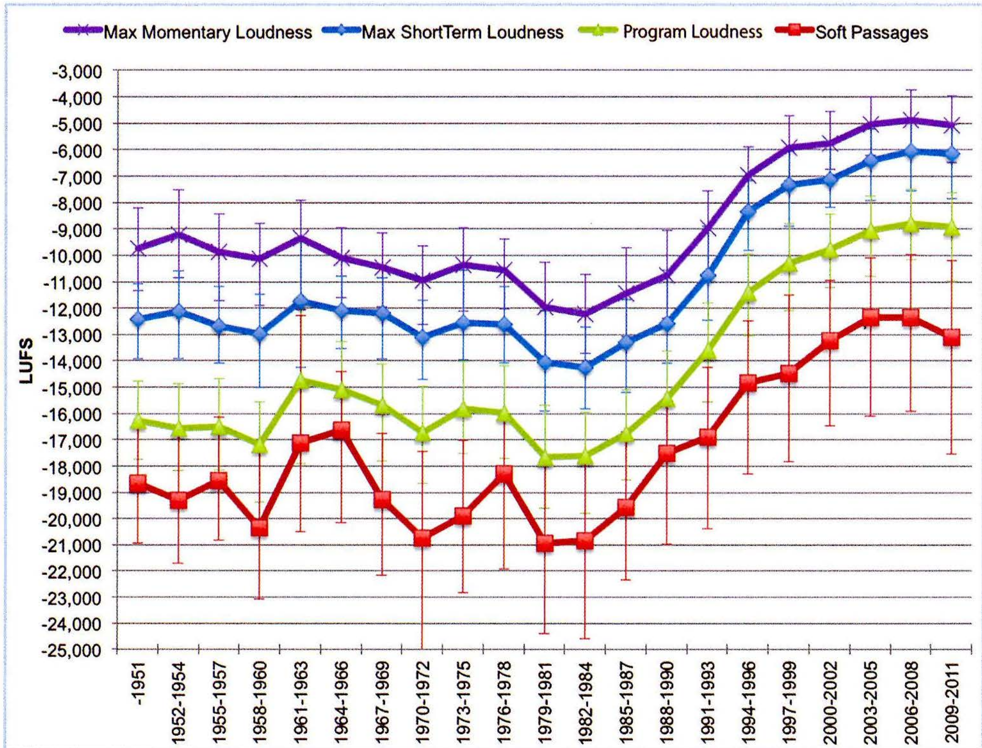


Increase in dynamics processing *required to achieve a given loudness, throughout a 30 year period. (Processing estimated by Katz, loudness data from Ortner).*

the level for their own comfort. This marks the end of the digital race. There is no headroom left to raise the average level. In the line of fire at the climax of the loudness war, I've produced a few recordings with the kind of processing shown near the top of this chart. Many of my clients accepted it in order to be competitive, but would have preferred to use less processing if not for the loudness difference between them and their competition.

How Low Did We Go?

Ortner's statistics extend from 1951 through 2011. It's interesting to observe the trend of soft passages versus program loudness (pictured below). It appears that mastering engineers may have reduced their tendency to raise soft passages, at least past 1994, judging by the distance between the red curve (soft passages) and the green curve (program loudness). Some have claimed that the introduction of car CD players and portable digital players fueled the loudness war. If that were so, we might see a rise in soft passages after the introduction of the car CD player in the late 80s or the iPod in 2001, which is not evident here. When soft passages are somewhat softer, the ear perceives a more "open" sound — the music breathes more. However, the graph is not a measure of the density of a program, and certainly a dense program (which is frequently loud) would be easier to accommodate to the needs of a noisy car or a jogger in Central Park. The solution to noisy playback environments, as I've mentioned, is to introduce noise-sensitive circuitry into portable and car players, not to overcompress the master recordings.



How Low Did We Go? *Engineers allow recordings to get soft, measured between 1951 and 2011. Notice the trend of soft passage levels is not quite as steep as that of the loudness, especially past 1994, which implies that we are allowing soft passages to get softer. (Graph produced by Ortner)*

How Low Is Too Low? Note that at the end of the loudness race, our compression practices became far more aggressive than what the consumer will tolerate, even in the car. The public will accept a dynamic range much greater than we have been allowing, according to this dynamic range tolerance chart (page 249, from research performed by TC Electronic). The chart is intended to be a guide, not a prescription. Furthermore, all media are not alike. For example, the consumer has a greater tolerance for music dynamics than for film or television dialogue, where intelligibility and room noise are greater issues. Primarily we need to be concerned about music dynamic range in the noisy venues, at the right of the chart, and should make judgments about specific songs, rather than pigeonholing every song into an arbitrary numeric limitation. If we avoid mastering to the lowest common denominator, then good dynamic range can be enjoyed in the environments on the left side of the chart. The best solution is venue-specific noise-sensitive gain-riding, like the Bose AutoPilot, which works very well in my car.

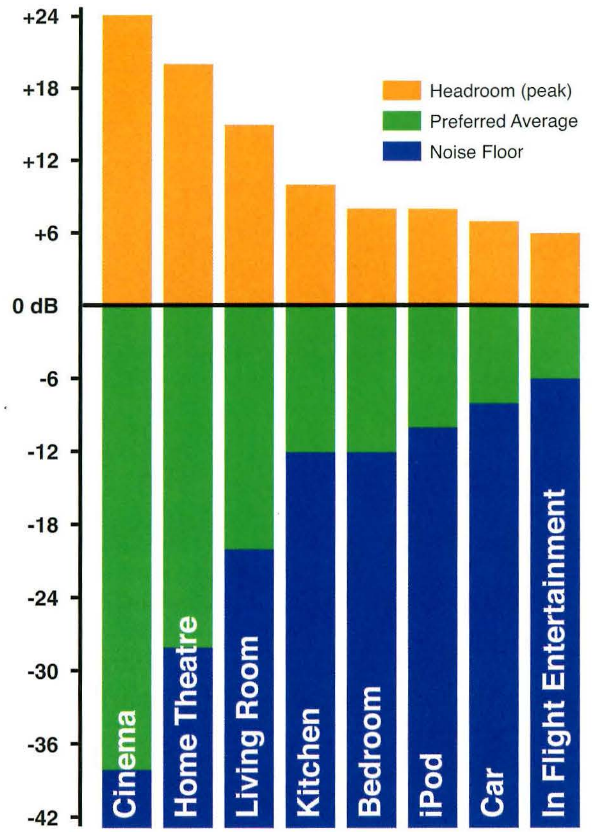
Evidence of Peak Limiters

Normal recordings do not need peak limiting unless the peak signal is purposely maximized to full scale and we wish to raise the average level even further. Ortner's research (see figure page 250) shows that as soon as the weapons became available, engineers began to employ them. The CD was introduced in 1980, but digital limiter plug-ins did not become readily available until the early 90's. A proliferation of peaks higher than -1 dBFS can only be produced by the use of peak limiters in mastering. The graph shows a dramatic increase in the number of samples per second higher than -1 dBFS right after 1991. Orange is the median, and the grey bars show the extremes. Note that as soon as limiters became widely available around 1991, the height of the grey bar increased, showing a greater range of extremes. This indicates that some engineers were more conservative, and others pushed the limiters much further, but as the years progressed, more engineers began to use peak limiters. Now it's considered "normal" to peak limit a pop CD, even though this is not a technical requirement.

II. Loudness Normalization Ends the War

History

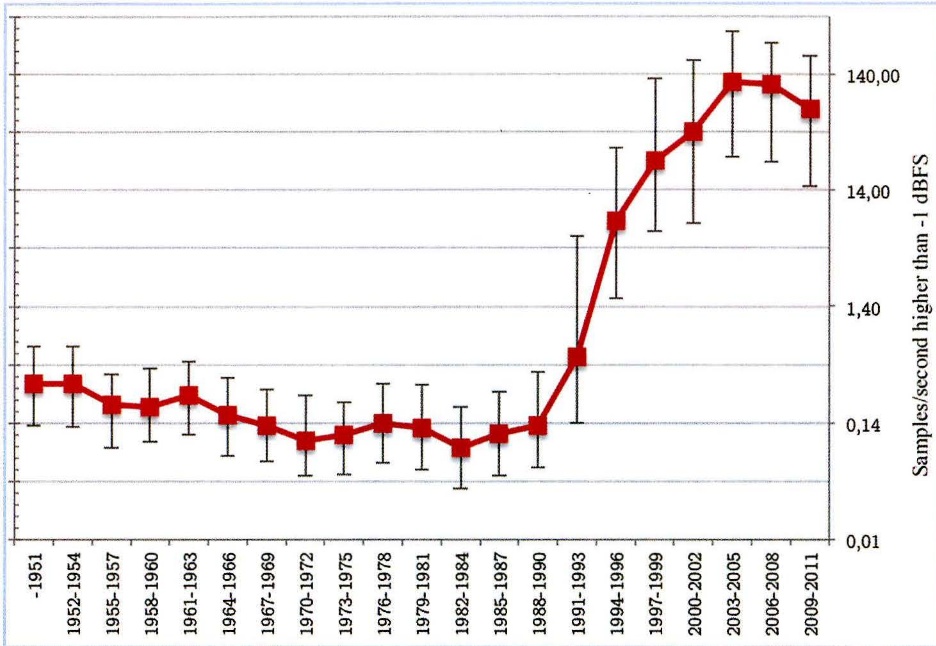
There have been a few attempts to suggest voluntary loudness normalization. Some of us tried to get our fellow mastering engineers to agree to pull levels back,



This dynamic range tolerance chart is intended to be a general guide for dynamic range concerns, not as a numeric prescription (see text). (Courtesy of TC Electronic)

but a voluntary approach simply doesn't work. This doesn't fix older recordings which have already raised the bar. The only solution is a global one, where the entire consumer experience is loudness-normalized, from old recordings to new ones. There are already many good indications of a global movement towards loudness normalization:

"Loudness is a drug!"



Evidence of peak limiting. See text (page 249). (Graph produced by Ortner)

- ATSC TV broadcast in the U.S. (-24 LUFS target)
- EBU broadcast throughout Europe, first in television and spreading rapidly to radio, with a -23 LUFS target
- iTunes Radio, with the Sound Check algorithm and approximately a -16.5 LUFS target
- Sound Check in iTunes file playback
- The Game Audio Initiative. Four game manufacturers have standardized on R-128, with a -23 LUFS target level, which has exceptional headroom to accommodate the sound effects of the game. Bravo!
- Numerous Internet streaming stations

Of course there are holdouts, most notably U.S. terrestrial radio and Satellite Radio with their egregious processing. The ATSC standard was inspired by the FCC's **CALM Act**, and we can only hope that similar legislation will be passed for U.S. terrestrial radio.

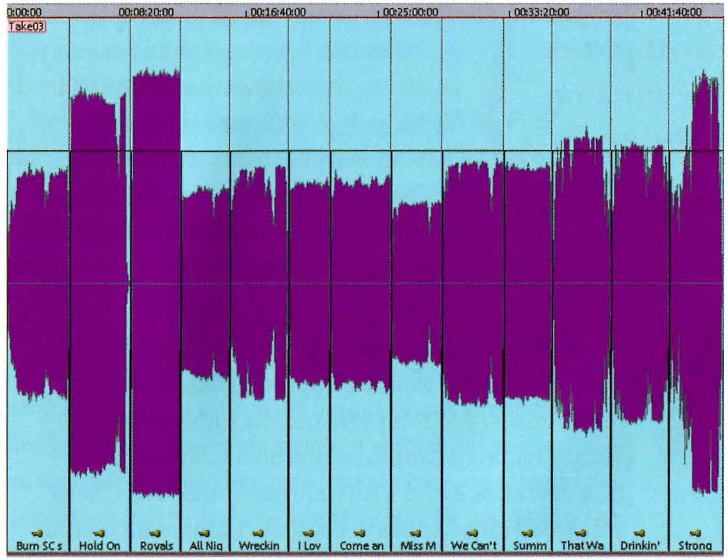
Pop or Flop

EBU-style loudness normalization is very different from the older compression and processing practices of terrestrial broadcasting, satellite radio and some stations on the Internet. EBU-style does not use any dynamics processing—normalization only adjusts the level. This has a very different effect on our appreciation of the sound of a recording, as Ian Shepherd emphasizes:

> Standing out from the competition is crucial, right? Which is why people try to make things loud. But if iTunes Radio decides to turn a "loud" song down further than expected – it will actually end up sounding quieter. FM radio processing would have made it sound just as loud, but more distorted – and many people wouldn't have noticed or cared. But they'll notice if it sounds quieter! The times, they are a-changin'... Dynamic is the new loud.[5]

In fact, loudness normalization exposes the wimpy nature of overcompressed recordings. Now the overcompressed songs sound as squashed as they always did, right from the first beat. Here is a waveform (page 251) from iTunes Radio's pop music channel. This appears very different from peak normalization, and from terrestrial broadcast—recordings with a more dynamic nature are permitted to reach higher peak amplitudes, and overcompressed recordings are brought down; they look (and sound) very "small." There is so much peak capability available (in the blue areas above and below the violet waveforms) that if an artist comes along who takes advantage of this and is even a little bit more dynamic than the rest of the crowd, he or she will lead the pack in clarity and impact on iTunes Radio!

Some naysayers have resisted the notion of loudness normalization, claiming that before normalization they could make their song louder. But this is a myth. Engineers have never had the power to affect how consumers play their songs. If you turn it up, they'll turn it down! It is true that before the invention of normalized media we could compress a song so it would gain momentary attention, before the consumer hit his volume control, but in this new world, everybody is on a level playing field. With loudness normalization, artists can choose to make recordings with any amount of compression, and they regain a greater range of expression. Normalization restores their freedom of choice to produce undistorted recordings and recordings with microdynamics, which they lost at the climax of the war (except for the most adventurous of artists who did not care about loudness competition and preferred to make recordings for their sonic art).



Waveform showing 13 tunes played on iTunes Radio's pop music station. Notice how song #8 has been brought down very much because it is so compressed that its program loudness is very high in comparison to the rest. Nevertheless, the vast majority of songs on the pop station exhibit rather squared waves. This will change as soon as pop producers catch on to Sound Check.

Prevent Another Tragedy

Of course iTunes Radio is not the only music broadcaster on the Internet, but it does implement a loudness normalization system with sufficient headroom to play the vast majority of recorded music tracks with their dynamic range intact. I hope that other Internet networks will follow suit, and lower their target levels down to iTunes' conservative value of -16.5 LUFS (approximate). This is important not only because of sound quality, but *we must prevent another loudness war in streaming media*. I hope the tragic lessons of this Chapter will be communicated: *Streaming services must unite, agree on a sensible, standardized target level. If not, most artists will seek the lowest common denominator and there will be another loudness war—forever.* Already consumers are complaining that some services appear to be "too low" while really the issue is the services which are too loud to preserve dynamic range or normalization. Loudness normalization can end the loudness war only if all streamers unite on a single target. Please visit the links to stay on top of this important issue.

Advantages of An Ecosystem

Apple is the first company with a complete ecosystem: computer-based-player, portable player, radio network and music store! Consumers and producers audition streaming music on the identical playback device and DAC that they use for playing their own music collection. For the first time ever, consumers can directly compare the sound quality of their own music collection with how that music sounds on the radio. A direct comparison was never possible

"We must prevent another loudness war in streaming media!"

> *"Pink Floyd plays as loud as P!nk on iTunes Radio."*
> — Ralph Kessler

with terrestrial radio; the CD player and portable player sound noticeably different than the output of FM and Satellite Radio. Loudness normalization is a boon for AAC encoding, because coded material sounds better when it is not saturated or its peaks approach full scale.

As introduced in Chapter 16, *Sound Check* is Apple's loudness normalization technology. Sound Check has always been available as an option in iTunes music preferences, and it is permanently implemented in iTunes Radio (it cannot be disabled). The introduction of iTunes Radio with a permanent Sound Check will accelerate Apple's decision to implement Sound Check as a default in iTunes file playback. Otherwise consumers will hear a big difference between songs they purchase and play on iTunes and how they sounded on iTunes Radio.

Here's what Apple has to say about Sound Check in their Mastered for iTunes document:

> The effect of Sound Check, as well as other volume-controlling technologies, is that songs that have been mastered to be too loud will be played back at a lower volume, letting listeners more easily notice any artifacts or unintentional distortion. Because many such technologies are available to listeners, you should always mix and master your tracks in a way that captures your intended sound regardless of playback volume.

How it Works

I believe that iTunes Radio broadcasts the identical AAC file that can be purchased from the iTunes store. The loudness metadata is embedded in this file, and interpreted by the computer radio player as a gain shift prior to playing the music. This mechanism is identical to iTunes file playback: iTunes reads the metadata and Sound Check adjusts the gain on playback. The only difference between the iTunes approach and the EBU approach (besides a different target level) is that European terrestrial radio adjusts all material prior to broadcast, stores the adjusted file, and broadcasts it at the intended level of the consumer (metadata is not used). In the U.S., the ATSC approach is to use the Dolby Dialnorm metadata in the AC-3 bitstream, so the consumer's TV makes the loudness adjustment. This is more like the iTunes radio approach. iTunes players support WAV and AIFF files, which do not incorporate loudness metadata, so the application keeps a separate metadata database. This approach is not as reliable: I have seen failure of normalization in some cases when the database gets lost or misplaced.

Obstacles to a Fully-normalized World

Disc playback cannot easily be loudness-normalized, because the system would first have to scan the disc to determine its loudness before it could be played. So when the consumer inserts a CD, it will probably sound louder than the same songs played on iTunes Radio. On iTunes with Sound Check engaged, he will hear a difference in level once he has ripped the CD's songs into iTunes. We have to explain these issues to our clients. However, Apple could use the Gracenote database to determine the identity of a CD (as it does) and know the loudness value of all the songs even before the consumer hits the play button. If this is not practical, iTunes could default to a standardized attenuation of, say, -6 dB and then slowly revert to an accurate target as soon as it finishes analyzing the whole CD.

Blu-Ray and DVD players are a lost cause. Disc producers have long ago defeated the efficacy of Dialnorm, Dolby's loudness normalization system, encoding a tricked-up loudness metadata value instead, so the player will bring up the level of their disc. It's the old "everything louder than everything else" philosophy. In contrast, Apple is in charge of the loudness metadata for songs purchased from the Apple store. It is impossible for an outside producer to defeat the system. That's why I can't wait till discs become obsolete and consumers only have files to play. The CD, DVD and Blu-Ray remain the only media that inspire producers to compete for a louder record. It will take a long time for discs to die, as many third-world nations have just gotten past cassettes!

III. Production Practices in a Loudness-Normalized World

While We're Waiting

It'll be hard to convince our clients not to smash their masters if they are producing pop masters for CD. Here are a few strategies for getting the best-sounding product in as many places as possible:

- Tell them that the CD is dead and has been replaced by digital downloads. Well, that strategy won't work yet, but it's worth a try!
- Explain that the portable media player is rapidly replacing the disc player in cars, because it's convenient and listeners can choose from their music collection instead of just six CDs. They can also play podcasts and Internet radio. Demonstrate the advantages of loudness normalization and explain how to turn on Sound Check. Explain that Sound Check will inevitably be turned on by default on their iPhone, iPod and computer. Tell them that many consumers have already turned on Sound Check because they like not having to ride their volume control. Tell them to spread the word about Sound Check to their own fans.
- Demonstrate iTunes Radio. Play the "Pure Pop" channel and show how the smashed hits sound poor compared to the ones on the "Hits of the 80s" channel. Explain that if you make their master the right way, it will shine over the competition on the "Pure Pop" (or any other) channel. Demonstrate how your master sounds better than the competition by downloading files from the iTunes store and playing them back-to-back with your master with Sound Check turned on. Explain that these same masters will sound very good on terrestrial radio because less processing means less distortion on the radio — but they will be just as loud or even louder than the competition, even if the CD of the competition sounds louder to them.
- Point out that AAC encoding cannot tolerate over-processing, and that it sounds noticeably better when levels are turned down even a little bit. Explain that even if you make their CD master as hot as the hottest competition, the AAC version should be turned down at least a little bit from the CD master.

Album Normalization vs. Singles Normalization

One difference between iTunes Radio and the EBU/ATSC approach to broadcasting is that standard broadcasters can treat a record album as a complete *program* or as a portion of a complete program. This is important in classical music, where the adagio movement of a symphony should be played softly compared to the allegro. To my knowledge iTunes Radio is not equipped for

"I want it as loud as everything else but I don't want to lose the dynamics or add any distortion."
— Contributed by Brian Lucey

> *"Loudness Normalization in streaming/broadcast has the potential to restore full dynamic range to pop record production."*
> —Alan Silverman

album normalization, which means that it can only stream individual movements, each adjusted to the target level. This is called **Singles Normalization**, also known as **Track Normalization**. Track normalization is a form of compression. For example:

Mastering engineers and producers make albums with a bit of dynamic range so that they flow well. Ballads are intentionally made softer than rockers; this is usually the artist's intent. Let's consider a playlist combining a Beatles album with a Sinatra album (not far fetched, if you have wide-ranging musical tastes). Remember that these albums have previously been peak-normalized by the mastering engineers. These are conceptual demonstrations, not actual measurements. Our Sinatra album consists of a singer with a big band (Figure A, page 255). Its Program Loudness (PL) is indicated by the dashed white line in the violet area. It has a bigger PLR than the rock album, as indicated by the taller yellow bar on Sinatra's loud song. That's why without normalization in the player, the Beatles loud song sounds louder than the Sinatra loud song. The loud song on the Sinatra album is only as loud as the soft song on the Beatles album. So when the two are played back-to-back without normalization, the consumer has to turn up his monitor gain when the Sinatra album comes on.

In Figure B, we see what happens when the normalizer uses track normalization: it fixes one problem and creates another. Track Normalization lowers all the loud songs to the target level and, if necessary, raises the soft songs to the same level. So now both albums sound compressed: the loud numbers cease to swing because the soft numbers come in too loud and ruin the intended contrast. This is the endemic problem with singles (per track) normalization.

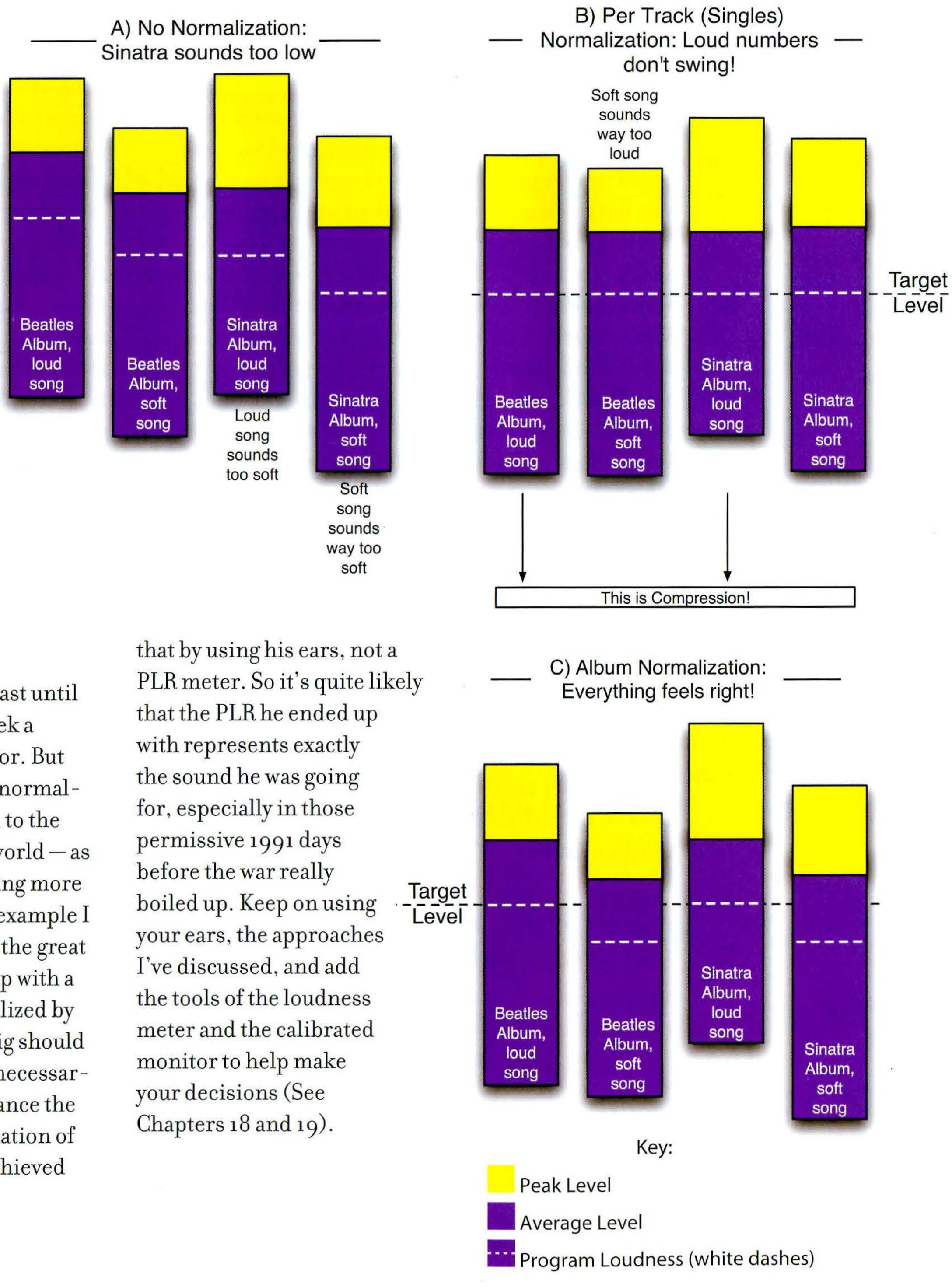
However, with album normalization: Everything feels right (Figure C). The loud song on the Beatles album and the Sinatra album are now equally loud, but the soft songs remain soft in the same proportion as on the original albums. Dynamic contrast is back—so the sound swings!

It's important that we learn how file-based playback and streaming services apply these different normalization techniques during streaming or home playback, so we can inform our clients. iTunes' file playback intelligently and transparently switches between album and singles mode, making the experience seamless and enjoyable for the end user.

Although I generally prefer album normalization, there is at least one advantage to singles normalization: the producer will not need to make a louder singles version of the ballad in the album. iTunes Radio will raise the ballad up automatically and iTunes file playback will singles-normalize automatically when appropriate. Frankly, even with terrestrial radio, it has usually not been necessary to make a loud single version of any song, because radio's compression has always brought up soft material.

Confessions of a Mastering Engineer

In an attended session, when confronted with an unhappy client, every engineer knows this secret move: turn up the volume control! "How's that?" "Oh, that sounds great, so much better. What did you do?" If we are honest with ourselves, the illusion of loudness is the most important tool that we have. More than half the battle is getting the loudness right. Loudness normalized media should cause all of us to question the assumptions and even the conclusions we've reached about our own work over the years. Hopefully, even better-sounding work will result. I've been working with 12 to 14 dB headroom for a long time, and many times with even less, but I'm certain that some of the styles of music I work with will benefit from more headroom and obtain a more open sound character. It will be liberating to have even a couple of dB more headroom to play with, to create a real crescendo without having to push it into a peak limiter.



What is A Good Sound to Aim For in a Loudness Normalized World?

12 to 14 dB PLR is probably the average of all recordings made since about 1950, at least until the loudness war heated up, so if you must seek a numeric guide, that's a good number to aim for. But truly, the best sound to aim for in a loudness normalized world is the same one that sounded good to the mastering engineer in the non-normalized world—as long as that sound would not benefit from using more of the available headroom. The Mellencamp example I cited at the head of this chapter, mastered by the great Bob Ludwig, has a PLR of 12.5 dB, so it ends up with a true peak level of about -4 dBFS when normalized by iTunes Radio. Does this mean that Bob Ludwig should remaster the album with a greater PLR? Not necessarily. He worked really hard to preserve or enhance the original mix, to achieve just the right combination of punch, warmth, clarity and impact, and he achieved that by using his ears, not a PLR meter. So it's quite likely that the PLR he ended up with represents exactly the sound he was going for, especially in those permissive 1991 days before the war really boiled up. Keep on using your ears, the approaches I've discussed, and add the tools of the loudness meter and the calibrated monitor to help make your decisions (See Chapters 18 and 19).

Peak Limiters

In 1980, digital sample-level peak limiters were not yet invented, and we didn't see much need for them. Limiters were primarily made to serve the loudness war. But in a loudness-normalized world with a sufficiently low target level, mastering engineers will hardly ever need digital peak limiters, except where a peak limiter helps obtain a particular sound character. Some of the current internet streaming services use too high a normalization target, and incorporate their own peak limiting to keep the sound from overloading. This distorts the peaks and changes the sound. However, iTunes' Sound Check, with a low target level of about -16.5 LUFS, accommodates the vast majority of recorded music without causing it to hit full scale peak, and preserves the music's recorded sound character.

I am concerned that iTunes allows a positive normalization gain, so the rare material with a PLR higher than about 16.5 dB could clip when normalized in iTunes or require a limiter. But album normalization mitigates the chances of clipping a soft tune as long as the loudest tune on the album has a PLR of 16.5 or less. I would rather see iTunes eliminate the possibility of positive normalization gain. We have to investigate that situation further.

Loudness normalization is becoming a part of our audio life. Since many consumers use iTunes for their music playback, we should encourage our clients to use iTunes to hear what the consumer hears, and to turn on Sound Check to hear how songs will sound on the radio as singles. They can switch back and forth between iTunes Radio and iTunes with Sound Check and know in confidence how well their music will compete with what's playing on the radio. When they listen to the album in album mode, the softer songs will play softer than in singles mode, so we have to show them that there is a difference. This complicates our process a bit, but look on the bright side: on track-normlized radio, ballads will sound just as loud as rockers, so the client can pick a single from any tune in the album.

"We have to stop giving the impression that we have any control over the consumer's volume control. In the end we are just making Muzak if we overcompress."

— Clete Baker

Moving On

This completes the first of three Chapters about *The Loudness Revolution.*[6] In Chapter 18 we will learn how to take maximum advantage of loudness meters.

1
A 40 foot (13 meters) recording horn was once built by Edison. But the longest practical acoustic recording horns were about 7 to 9 feet long. All this was superseded by the invention of electrical recording in 1925.

2
The late Gabe Wiener's classical recordings noted the SPL of a short passage, encouraging listeners to reproduce the "natural" SPL of the ensemble. I used to second-guess Wiener by adjusting monitor gain by ear, then checking against his listed level. Each time, my monitor gain was within 1 dB of Wiener's recommendation.

3
Composers have a similar dilemma with the acoustic advantage. They mark chamber music scores with the same dynamic markings as full orchestras because a forte mark is for the benefit of the individual player as well as the ensemble. But forte in the full orchestra comes out much louder. And it sounds wrong to reproduce the chamber group at the same loudness as the symphony orchestra.

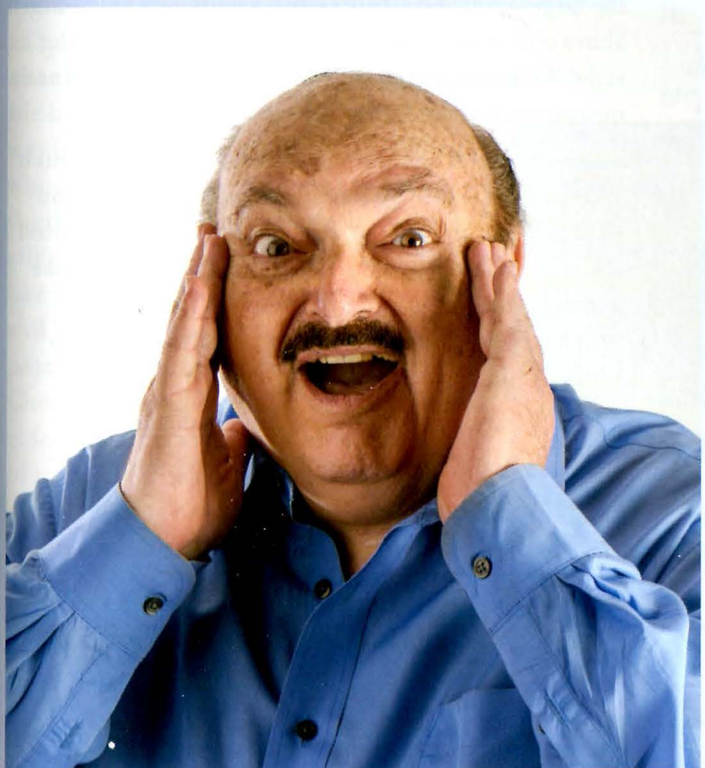
4
Ortner, Matthias Rudolf (2012) Je lauter desto bumm! (The Evolution of Loud). Master's thesis, Danube University Krems

5
Shepherd, Ian, (Mar. 12, 2014), http://productionadvice.co.uk/u2-21st-century-loudness-secret/

6
I credit Florian Camerer for coining the term *Loudness Revolution.*

CHAPTER 18

The Loudness Revolution: Loudness Metering: It's Time!



Introduction

Excuse me for rocking the boat, but now is the time for all mastering engineers to replace our meters with loudness metering. Recording and mixing engineers should also consider this change.

I. Using Loudness Meters

Peak Metering is Gone (and that's a good thing)

The first thing to notice about the new EBU-mode loudness meters is that they do not display the active (moving) peak level — and they should not! This is to discourage users from being tempted to perform *peak normalization*, that is, watching the peak level and trying to place it at full scale. It bears repeating that the ear does not judge loudness by the peak level. We want to keep those pesky peak meters out of the eyes of producers, musicians, and engineers! The transition will take time, but we have to start. The only peak indicator on an EBU-mode loudness meter should be a True Peak warning, in case the true peak goes over a predefined threshold. For EBU broadcast this warning should be set to -1 dBTP, for ATSC to -2 dBTP. Unfortunately, -2 dBTP robs precious headroom, so if necessary, reduce program loudness to -24 LUFS to regain the peak headroom and avoid peak limiting. For music productions working at a higher target (e.g. -16.5 LUFS), I still recommend setting the peak warning indicator threshold to -1 dBTP, because the peak level will increase after lossy coding (to be described below). If the music is later broadcast, when it is loudness-normalized, its loudness will be reduced, and PLR will be retained, as illustrated in the figure on page 223.

Using the Moving M Meter or S Meter

In practice, while producing a music program we need to have a "moving-needle" meter that gives us a

> "For o LU= Forte, set the M Meter's o LU point to ~2 or 3 dB above the target loudness."

general idea of the current program activity and helps us achieve the desired target program loudness. Many engineers are accustomed to using a meter where a 0 value represents *forte*, meaning +2 or +3 is approximately *double-forte*, and -2 or -3 is *mezzo-forte*. In other words, we adjust our monitor gains so that forte sounds loud enough to our ears, and coordinate the meter calibration so the meter reads *0 LU* at that value. This is called a "center of gravity" approach. For this purpose I recommend the **M** (momentary) meter, and that we abandon the VU and all previous non-weighted meters. To review, the *momentary loudness,* M for short, is the loudness that you hear *now.* **M** is averaged over a 400 ms period, which corresponds well with the VU meters that many of us are used to, and with our short term perception.

The other EBU scale is **Short-term loudness**, abbreviated **S**, with a time window of 3 seconds. It can be used as an indication of long term tendencies. Some engineers are finding that S correlates well with the intended integrated (program) loudness, even for dynamic material. Calibrate 0 LU on **S** to the intended target. Being an old-time guy with forte in my head, I may not get used to S, but some of you may prefer it. Regardless whether you prefer M or S, glance at the accumulating integrated loudness (program loudness) to get an idea how the program is coming along.

Metering for A Loudness-Normalized Medium

When mastering for a loudness-normalized medium, if we use the M meter as a *forte* meter, the way we operated VU meters back in the day, more of the program will fall below forte than above it, so the average loudness will probably end up at *mezzo forte*, which typically falls 2 or 3 dB below forte. *Try calibrating the M meter's 0 LU point to 2 or 3 dB above the destination target* — for example, approximately -13 LUFS for an iTunes target. When we do, we will probably find that the full program ends up "on the money" or very close to it. Regardless, experiment, and you will soon find a calibration whose M meter movement you like and yields the desired target loudness.

An *integrated* PL means that there are some levels above and some levels below the average, so another approach is to use the S meter with 0 LU calibrated at the target level and aim for equal excursions below and above 0 LU — whichever approach suits your working style. We are all united with a target integrated loudness measured in a standard way.

The M meter seems to favor dynamic music, which suits me just fine! I've found with more dynamic music, after monitoring with the M meter, the integrated PL tends to fall below the target. This occurs because dynamic music has more frequent soft passages, which lowers the average. There are several ways to address this:

- Recalibrate 0 LU to a higher LUFS value, or
- Raise the level of the resulting file to yield the desired target, or
- Remaster by raising the level of some of the loud passages, thereby increasing dynamic range even more, but also bringing the master up to target level, or
- Reassess and remaster by raising the level of some of the extra-soft passages, which lowers the dynamic range, but still might be a good thing or
- Try using the S meter for our next program reading

Before sending a low-level program to a distributor or network, consider that it may be returned to be re-mastered at a higher level. At least inspect its true peak level to ensure that if it is to be raised, the true peak level will not exceed the permitted maximum!

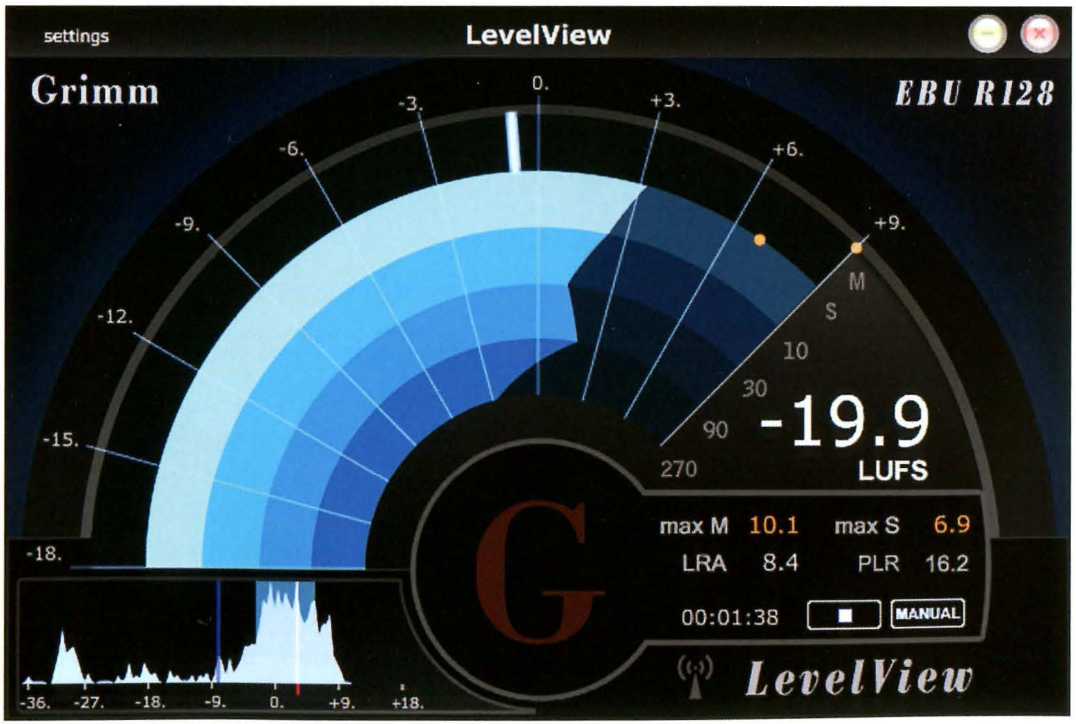
Let's suppose we are producing music, and we want it to sound the best it can on iTunes Radio. We may want to make the loudest song on the album be about -16 LUFS, so that iTunes Radio will not alter its level. Let's say our great-sounding song ends up with an integrated loudness of -16 LUFS, and a PLR of 12 dB, which is not at all unusual. In this case, there is 4 dB unused peak space between the program's highest peak level and digital full scale, illustrated in the first figure on page 223. Let's say 3 dB to allow for peak level to increase after encoding to AAC or mp3. Raising the level of the file by 3 dB will not change the sound in any way (if dithered), except that couch potatoes who never touch their monitor gains will think it is louder. This is certainly permissible for a current competitive medium like CD; nevertheless, iTunes Sound Check will lower the level right back down. Couch potatoes will already have a preferred Sound Check monitor gain. Since we already decided the program sounds good with a PLR of 12 dB, there is no harm in raising it 3 dB and delivering this hotter program to Apple, as long as all parties realize it will sound lower after Sound Check gets hold of it. As mastering engineers we want to ensure that our client doesn't come back and complain that their song sounds lower than the master on the radio. This is why I recommend that all iTunes users turn on Sound Check so as to hear what it will sound like on iTunes radio.

Don't Be A Constant Meter Reader!

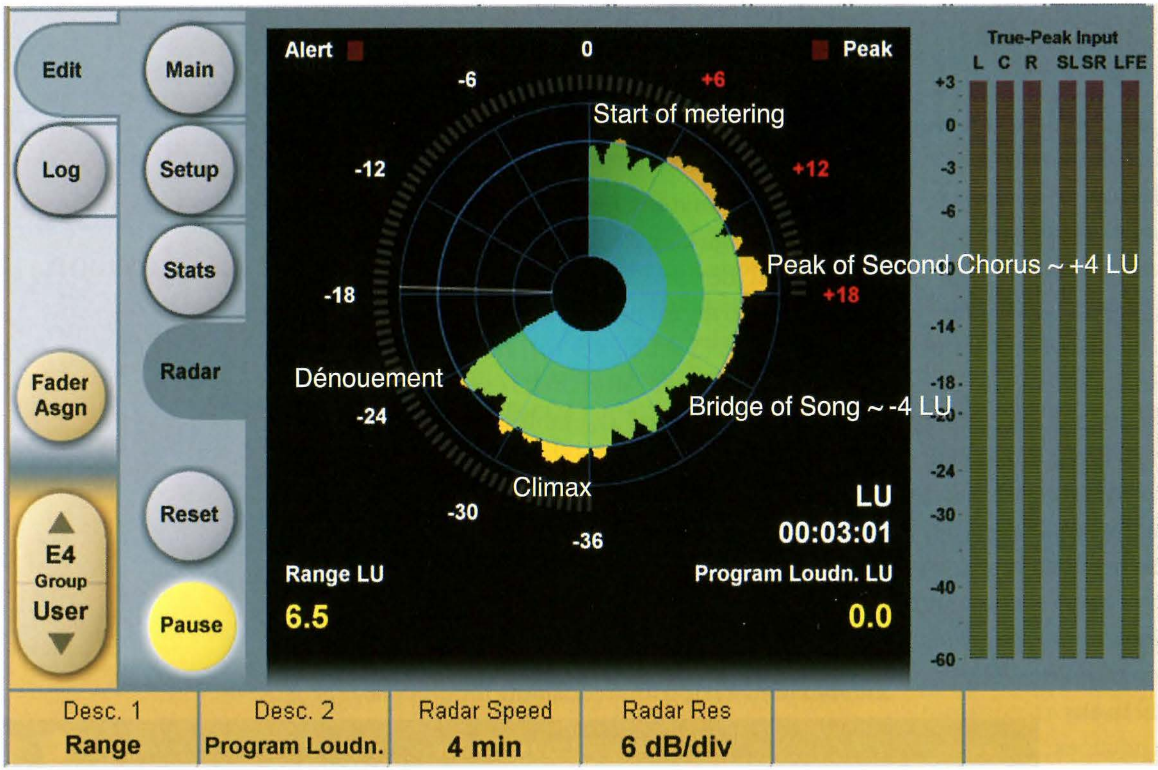
When I'm working on an album I'm not looking at the meter all the time. Generally I work on the loudest song first, with my monitor control set to a predetermined position (See Chapter 19), glancing occasionally at the loudness meter. After that, I essentially stop looking at the meter as I assemble the album. The rest of the album comes together almost completely by ear. Then I measure the album's integrated program loudness and true peak level and decide if it will work well on the network or destination medium.

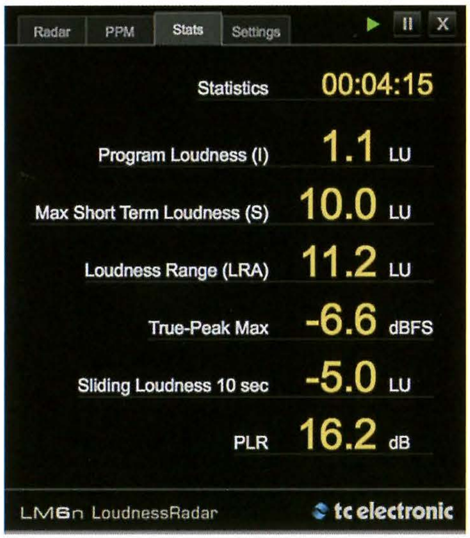
II. Three Popular Loudness Meters

The *Grimm LevelView* (below) includes a large number of metering time constants for every taste. For



The Grimm LevelView Loudness Meter





Top: TC Electronic Radar Loudness meter. The annotations in white are not part of the meter: I have added them to illustrate how the meter illustrates song structure. Bottom: Meter statistics page.

music production I'd rather see one time constant at a time, but you may like seeing them all. At the outside of the half-circle is the M meter (400 ms), followed by the S (3 seconds), then 10, 30, 90 and 270 second integration times. A histogram is shown in the lower left, where in this view you can see that the majority of the example material spends its time between -6 and +9 LU. LRA is indicated in the highlighed area on the histogram. Integrated, gated loudness is the large white numeric on the right side. The 0 LU point is adjustable to accomodate any conceivable target. The Grimm manual states, "Since the target at the end of the program is 0 LU, one should modulate approximately an equal amount above 0 LU as below it." This may work for the S meter, but the M meter will tend to ride above that.

The outer part of the *TC Electronic Radar Loudness Meter* (pictured at left) is a circle with a moving M meter, marked from -36 to +18 LU, or to +9 LU, adjustable in the preferences. The 0 LU value is fully adjustable, as is the threshold of the true peak warning indicator, so any conceivable target or standard can be accomodated. The inner part of the meter contains a powerful "radar" history which helps the engineer see the structure of the program at a glance. I've added annotations in white to this image to demonstrate. In this case I've set each circular division to 6 dB. Loudness between -6 LU and 0 are presented in light green. The loudest passage of this song reaches about +4 LU (in yellow). We can see that the loudness has never ventured below about -4 LU for the duration of the song, and the LRA is 6.5 LU (the statistical difference between softest and loudest passages). This meter is simple and fun to read even though it presents a comprehensive amount of information. The plug-in version can do faster-than-realtime measurements. The hardware and software versions can display PLR on the stats page (pictured bottom left).

The *Toneboosters EBU Loudness Meter* includes Integrated Loudness, LRA, current Maximum True Peak, PLR, an M and an S meter. It also displays true peak levels of individual channels, but as I mentioned, ideal loudness meters should display peak level only as a warning indicator. The user needs to concentrate on loudness, not the peak level. The 0 LU point is adjustable with -20, -14 and -12 LUFS values, plus the fixed ATSC and EBU scales, which should yield all popular target levels within a dB. These values effectively give the engineer a K-System meter that has been brought up to modern loudness measurement standards.[1]

Other Loudness Meters

Since the ITU/EBU specifications are a standard, a plethora of loudness meters have appeared, some of them freeware. The *Nugen* meter is popular and comprehensive. Channel D's *AudioLeak* loudness meter looks interesting, with histogram and logging facilities.

Loudness Meters vs. K-System Meters

I consider the K-System of metering to be obsolete, superseded by loudness meters with a few essential features. The most important thing is to get rid of the moving peak meter, which as I've said encourages engineers to peak to full scale. The terms "K-14" or "K-20" do not give a complete picture of a recording. Is that the PLR or the loudness? The fact is, we need to know *two parameters* to properly describe the average and peak characteristics of any recording: its program level and its PLR. These numbers can now be supplied by a few existing loudness meters, and I hope more to come.

Loudness meters that can replace the K-System meters must contain three essential features:

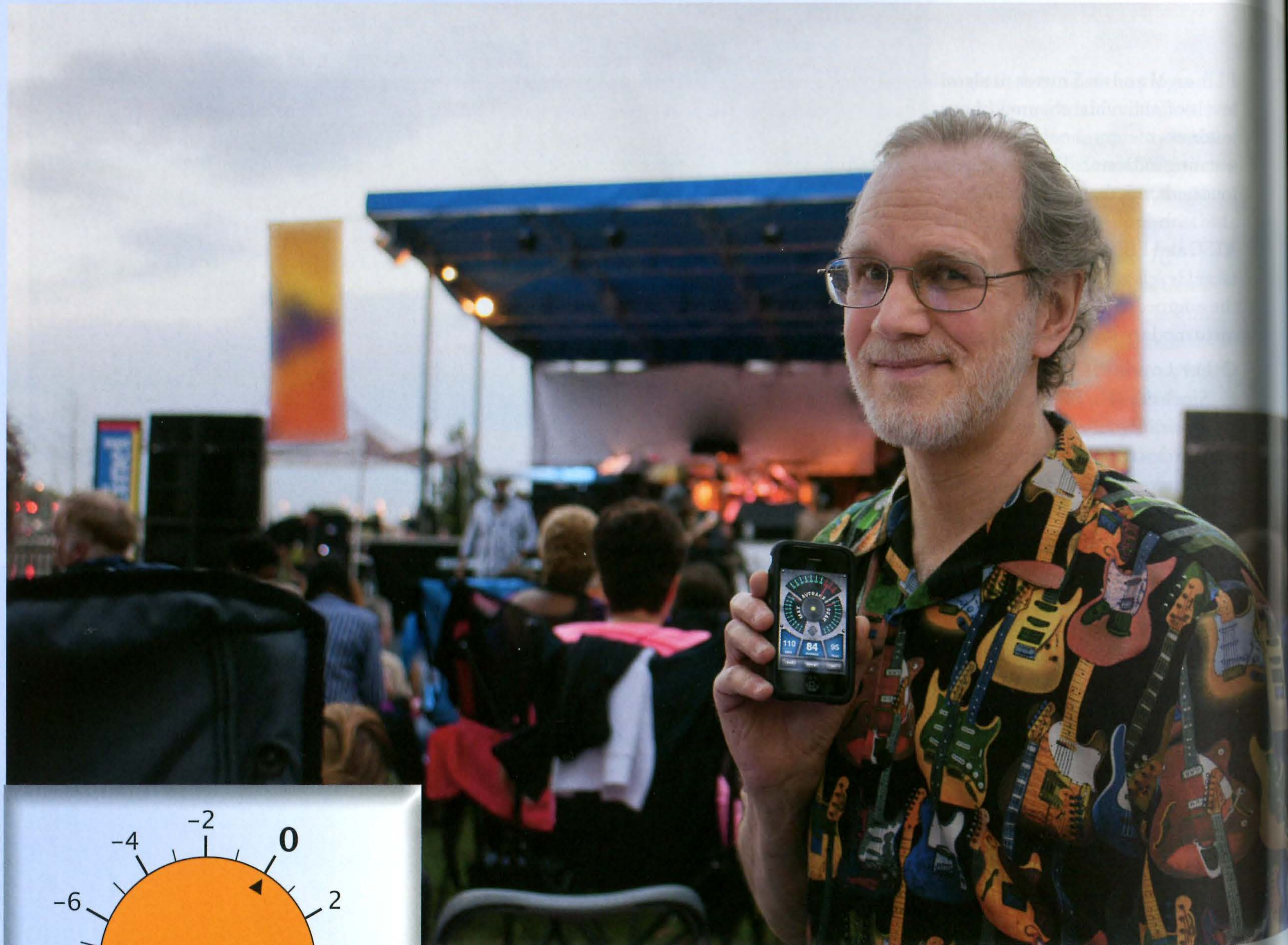
- Variable LU calibration. For example, -16 LUFS, -23 LUFS or any other value can be set to 0 LU at the user's demand.
- Hidden peak meter. Only a true peak warning indicator is available on screen. In some units the user can choose to see the moving peak meter, but it should not be a default.
- PLR indication.

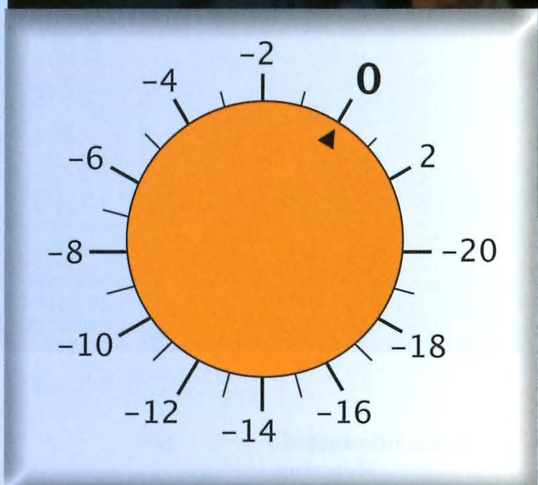
Metering and Monitor Gain

Meters should work hand-in-hand with monitor gain, which is why I've always advocated using a calibrated monitor control with the K-System. The good news is that calibrated monitoring is even more important in a loudness-normalized world, and we can update calibrated monitoring to involve R-128 loudness measurement. The next Chapter completes our Loudness Revolution trilogy with an updated look at calibrated monitoring for music.

1 Surround meters and vector displays are beyond the current scope of this book.

Toneboosters EBU loudness meter





Monitor Controller calibrated in decibels

CHAPTER 19

The Loudness Revolution: Calibrated Monitoring

I. The Whys and Hows of Calibrated Monitoring

What is a Calibrated Monitor System?

The motion picture theater already has a calibrated reference level, specified in SMPTE standard RP200, which has led to consistency in motion picture sound levels around the globe and plenty of headroom for sound effects and clarity of transients. But no such exacting standard exists in the music world.[1] Regardless, by calibrating our monitor controls in a standardized way, we will gain several advantages:

- The monitor gain is repeatable. We can return to our work tomorrow or next month in this room or in a similarly-calibrated room anywhere in the world, and play it at the identical monitor gain.
- Our work becomes more consistent, and we can produce mixes that will perform better together when assembled at the mastering house, even if different mixing engineers work on the same production, because we all speak the same monitoring "language."
- We can judge the program loudness and quality by viewing the monitor control position (to be explained).
- We can judge how much of the media's headroom we are using, or if we are running into a possible over-compression situation (to be explained).

The calibrated monitor control is like a water faucet with numbers around its knob: If we have to open the faucet very far to get a given water pressure out, then the incoming pressure must be low, and vice versa. In the case of audio, the more we have to attenuate our monitor gain to get the same loudness at our ears, the higher the incoming program level must be. These ideas are intuitively obvious, but what might not be so obvious is how powerful this control becomes when it is calibrated using these two precepts:

1) The mastering engineer likes to play forte passages at a consistent loudness.
2) When the PLR (Peak to loudness ratio) of the *incoming* program material exceeds the absolute value of the desired *outgoing* program loudness (PL), he must apply program compression to keep the *outgoing* peak levels from exceeding full scale.

The Calibration Procedure

We begin with a control marked in decibels (illustrated on page 262):

Next, we calibrate our monitor controller to produce a specified SPL, as measured with a test signal, with a sound level meter located at the listening position. The calibration signal is a special, narrow-band *single-channel* pink noise signal. Since an EBU loudness meter *combines channels* and is calibrated to read LUFS with a two-channel signal, the single-channel calibration signal reads -23 LUFS, which is 3 dB lower than our stereo goal. To repeat: our single-channel test signal reads -23 LUFS. Two channels at that level will produce -20 LUFS. Try this test with a loudness meter to get the picture: Send a -20 dBFS sine wave test tone to one channel. It will read -23 LUFS on the loudness meter. A stereo 1 kHz -20 dBFS test tone will read -20 LUFS on the loudness meter (channels are combined according to their energy, not their voltage). In multichannel, the combined loudness of all channels increases with the number of channels being played, as it would in a real listening situation.

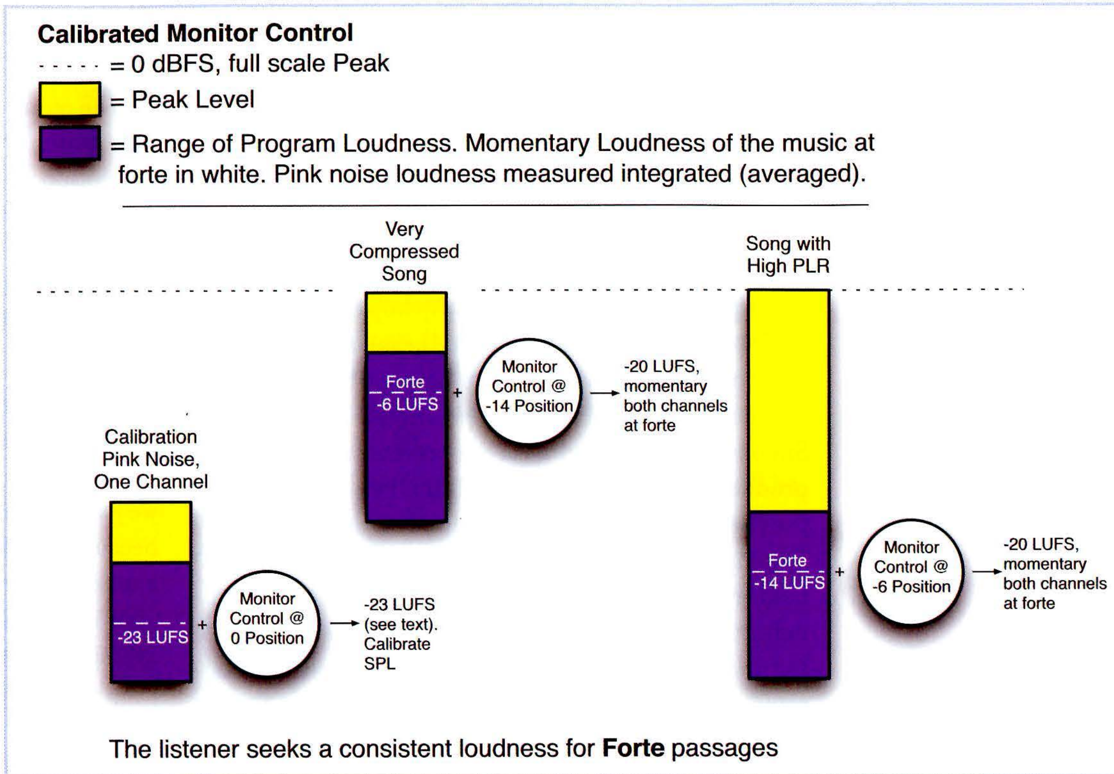
Then we adjust this single channel of pink noise to produce 83 dB SPL, C-weighted, slow meter setting, with the SPL meter located in the listening position. 83 dB (single channel), or about 86 dB (stereo combined using uncorrelated pink noise) is my interpretation of *forte* in a well-engineered mastering room. I am encouraging the use of the 83 dB calibration for a few reasons:

- Monitoring music at a slightly louder level encourages the engineer to produce a wider dynamic range. Engineers working at a greatly reduced SPL tend to overcompress their masters. It is still important to check the material at various levels, but I recommend returning to a higher monitor position for the actual mastering process.
- 83 dB is consistent with the SMPTE film standard, which has produced many magic dynamic and impacting sound tracks. It's especially useful for surround music productions.
- 83 dB is the most linear point on the ear's loudness contours. It enables the flattest reproduction and translates to the widest variety of user levels and venues.
- 83 dB SPL is the *reference SPL* at the 0 dB position of the control. The guides in this Chapter will help you find a good attenuation amount for a particular purpose. The 0 dB monitor position, which is very high, is probably not useful for producing most peak-normalized pop music, but I have used 0 dB to create music for audiophile or attentive listeners or wide range classical music.
- Setting 0 dB this high provides enough gain for studios with digital monitor level controls (which usually can attenuate but not boost) to audition soft material during production.

However, there are some potential issues with the 83 dB point:

- Only the finest, suitably large rooms with high-headroom monitors located far enough from the listener (e.g. 7 feet/2.1 meters) can support this loud a forte, so your needs may vary. For example, I checked a friend's mid-grade home theatre system, and it was clear that it could not support full calibration level and sounded best calibrating to 77 dB instead. Keep in mind his was a consumer playback system, not a studio monitoring system meant for creating programs.
- The smaller the room volume, the greater the perceived loudness for a given SPL. This is a psychoacoustic reason: Smaller rooms have less reverberation time, and the closer the loudspeakers the greater the proportion of direct sound. I'm rarely able to reproduce motion picture soundtracks at the same SPL as in the large theatre without them sounding much too loud. Still, I maintain the calibration as a reference point on the control. However, I can often reproduce compressed popular music forte passages at 83-86 dB in this pristine listening room when I want to listen loud. The ATSC[3] has provided a recommended sound pressure level for various room

volumes in standard RP-A/85 (See links). However, A/85 is intended for speech and music programs for digital television — I believe it is too conservative for producers mastering dynamic music intended for serious listeners. Keep RP-A/85 in mind, however, when producing programs intended for television or radio listening. Simply use the appropriate attenuation on the monitor control; I recommend -6 to -9 dB for producing television sound tracks in small rooms. Now you can see why I'm so enthusiastic about having a calibrated monitor control. We'll discuss A/85 later in this Chapter.



Calibrated Monitoring enables judgment of program quality and much more. Program with low or high level reproduces at the same loudness.

Proving the Concept

To demonstrate the power of calibrated monitoring, let's study this figure (at right). The principle is that we seek a consistent loudness for forte at the *output* of the monitor control. We start with the control set to the 0 dB position to play the calibration pink noise (left side of figure).[2] This yields -23 LUFS loudness *with one channel playing*, which will come up to -20 LUFS *with both channels playing* at equal level. Next we play a very compressed song (middle of figure) with a low PLR and a momentary loudness of -6 LUFS at forte, which is very loud. Think "there's high pressure coming into the faucet." To keep this song from playing too loudly, we must lower the monitor control to the -14 position. The net result: -20 LUFS loudness coming out of the monitor control. The second song is an "audiophile" recording with a high PLR and a momentary loudness of -14 LUFS at forte. To properly play the forte passages of this song, we must raise the monitor control up to the -6 position, also yielding -20 LUFS loudness coming out of the monitor control. This reveals the simple formula that the *sum of the momentary loudness and the monitor control position should be -20 LUFS* (our desired forte). This formula is the principle of the new loudness-based K-System of calibrated monitoring.

Now let's reverse the situation. Instead of playing back existing recordings, suppose we are mastering a piece of acoustic jazz which has been well recorded, and we want to produce a master which sounds open, dynamic, and clear, and will likely have a high PLR.

"The monitor control becomes a diagnostic tool, not just a knob that we turn."

In this case we can *preadjust* our monitor control to the -6 position, and master without meters, with our eyes closed. If we have learned our room and monitors, the program loudness and PLR will come out as predicted. But what if our hip hop client wants a competitive level because the loudness war is still going on (from his point of view)? We can start by placing our monitor control at the -15 position (I know where Snoop Lion lives), and work our processors by ear to produce our competitive recording. If we can convince the client to make a slightly lower-level master, then by simply turning the monitor control up to -13 or -14 before we start mastering, we can get better bass drum definition, punch, and clarity. This is because turning up the monitor lets us back down on the processing required to achieve the same forte loudness. There is a lot more to sound quality than just PLR, but regardless, the simple expedient of *presetting our monitor gain before mastering* is an important tool that will speed up our work. If we are finding that a master in the making sounds too compressed, the best solution is to turn up the monitor gain slightly and turn down the processing. The method allows us to return repeatedly to the monitor control position that has previously helped us to produce a particular style of music.

Of course we should check sound levels and quality at different monitor control positions — for instance to see how a program sounds at low levels — but I do recommend returning to the predetermined position to yield a more consistent, higher-quality sound. The monitor control becomes a diagnostic tool, not just a knob that we turn to get the right loudness.

The monitor 0 position corresponds closely to a standardized SMPTE playback level, so to play a theatrical motion picture from Blu-Ray, start with the monitor control set somewhere between 0 and -6 dB. As I have said, usually a motion picture sounds too loud at the 0 position because it was engineered for a much larger listening room.

The principle (which was not intuitively obvious) is that turning down the monitor control leads to needing more compression, and we can quantify that judgment to a great degree. The competitive recording we just mastered will be strongly compressed simply because we turned down the monitor control very far and made our fortes sound loud. In the last century, we approached monitoring empirically: as we raised the average recorded level, we turned down the monitor to keep our ears from overloading. *In the 21st century, first set the monitor control to a given position, then start mixing or mastering.* Of course the human ear isn't this precise or repeatable, but I wager that we are within about 2 dB in a good room.

The Point of No Return

I call the -9 dB monitor position the *point of no return*, the point below which audible degradation of the transient peaks and transient sound quality will occur (if we are trying to produce a program that combines good clean transients with good punch). Not coincidentally, this point corresponds with about 11-14 dB PLR, characteristic of the median of recorded popular music produced before about 1995, as we learned in Chapter 17. Many of us have produced programs more compressed than this, but we can use this -9 dB position as an indicator that perhaps we are overcompressing our material. With enlightened clients, we try to "back off" our processing, even a little bit, and if we are approaching the point of no return, 1 dB less PL can make a big sonic improvement.

This principle also means that if we master with a high monitor control position, we can produce music with a high PLR (crest factor). In fact, if we set the monitor control to 0 dB, we can liberate ourselves from all metering and simply produce the music, because our program will never overload! This is because the PLR of common mixed music rarely exceeds 14 dB, and only in extreme cases, 17 to 20 dB.

I advocate using a forte for music production that's loud enough to ensure making a dynamic and interesting-sounding program, but not so dynamic that even attentive listeners will need to ride their volume controls during the playback. Too much dynamic range is as bad as too little. In my room, forte is about 86 dB SPL (2 channels). If you prefer working at a lower level, run the control a little lower. In that case, your "point of no return" will fall below the -9 point on the control. However, if your monitors are telling you that fortes of nominally-compressed music sound much too loud at, say, 80 dB SPL, then your monitors must either be physically too close, or have too little headroom and are distorting. Small monitors and amplifiers self-compress. How can you tell how much compression to apply to the program if the monitor itself is compressing?

Closely-placed monitors exaggerate microdynamics: they give a skewed impression of transients, and this leads to overcompression, as does turning down their level due to their proximity. Nearfield monitoring also exaggerates the stereo spread, giving the impression the program is wider than it really is.

Monitoring Requirements for Professional Mastering with Adequate SPL

Very simply, if the monitor headroom is inadequate, the room is too small, or the monitors are too close, then 83 dB per-channel calibration sounds too loud because of the distortion in the monitors. A good mastering room should be long enough in the long dimension, at least 18 feet (5.5 meters), preferably as much as 30 feet (9.1 meters) to achieve a more even bass response. Loudspeakers should be placed between about 6 and 8 feet (about 2.1 meters) from the listener to approximate a living room situation. Amplifiers and loudspeakers should have sufficient headroom and low harmonic distortion, which will be discussed in more detail in Chapter 21. This means the monitors will not self-compress so they can reveal the compression or the microdynamics of the music. Under these conditions, a forte level of 83 dB (per channel, about 86 dB with two channels playing) sounds about right for pop music.

Calibrated Monitor Control with Loudness-Normalized Programs

The CD is not a normalized medium, and so active competition for loudness is still going on in competitive music genres. As we have seen, a lowered monitor gain leads us to overprocess. But when creating a program for a loudness-normalized medium, we have a lot more freedom of expression. We can still produce compressed programs if that's the style we're working in, but there is enough headroom to produce any style of

> *"-8 to -9 dB is the average monitor position for peak-normalized recordings made between 1900 and about 1995."*

recording (provided that the normalized system has a sufficiently low target level, like iTunes Sound Check).

If we're producing a music master, and want it to sound its best on normalized iTunes Radio, we should preset our monitor control to a position that we already know produces our desired forte loudness on iTunes Radio. I suggest between -7 and -9 dB and go for a loud forte if the program sounds good and doesn't exceed 16 dB PLR. This will help make a lively-sounding song that lands in the target loudness.

Producing mixed-media programs for television and radio broadcast. Here is a clear justification for using a lower SPL calibration for program production, since dramatic programs with speech and music played on television and radio are typically auditioned at a lower level. ATSC A/85 recommends reducing the playback level by as much as 9 dB in a small room with less than 1500 cubic feet volume (42.5 cubic meters). This would mean 77 dB forte, both channels playing. Or, about 74 dB at mezzo forte, which I think is a bit low and will result in overcompressed programs. I still hear a lot of overcompressed sound on U.S. television. I recommend setting the monitor position to -6 dB, and adjust your internal sense of forte to a lower SPL. This will keep us from producing a recording with too much dynamic range for bedroom listeners, but enough range to have some impact for the frightening moments of suspense programs and to allow music to be louder than speech when desired. Get to be known as a dynamic television mixer, not someone who squashes programs! If you end up with an integrated program level of the recommended -23 LUFS, then you have succeeded in recalibrating your whole system, including your ears. It doesn't take much time to get used to a lower SPL, but switching mental gears back and forth between television and music production does require some practice.

"With a monitor control position of 0 dB, we can mix with our eyes closed without fear of overloading peaks!"

A Holistic Approach

The calibrated monitor control is part of a holistic approach to gain structure. Someday every sound system will have a calibrated monitor, and everyone will speak the same language. Imagine a sound system operator who truly understands the gain structure of his system. You bring a soundtrack to play for your presentation, but there has been no time for a rehearsal or even to set levels. You tell the operator that your soundtrack is loudness normalized to -23 LUFS. With skill, the operator will know where to set his fader and get the right level in advance, without playing a note! This trick can be pulled off today, but it's amazing how few sound system operators think about calibrating their levels for this purpose.

Mixing With The Calibrated Monitor

Mix engineers can also benefit from a calibrated monitor. Here are some tips, including some related general good advice:

- Using a higher monitor position during mixing encourages making a recording with good clean transients. For example, preset your monitor control from 0 dB to no lower than -6 dB, which will produce a clean mix that falls in line with the vast majority and has acceptable dynamic range for home and car listening. If must mix using nearfields, you will probably need to turn your monitor control lower than that, but exercise the precautions mentioned elsewhere in this Chapter.

- If the mix overloads the peak meter, the monitor was probably set too low: turn the mix faders down, the monitor control up by the same amount, and keep on working.
- Do not use bus processing during mixing that's designed specifically to add "loudness": leave the loudness issues to the mastering session. This also prevents the vicious circle wherein a loud mix arrives at the mastering house, and we want to avoid having the client say "the master sounds lower than the mix."
- Never use a peak limiter on a mix to "protect" metered level, because there is no need to protect when you are already in charge of your own level. Of course use an overall peak limiter if it helps produce the sound you like, but I advise consulting with the mastering engineer if the program is to be mastered.
- Learn to use subtractive mixing. Avoid the practice of turning your monitor down as your recorded level creeps up. Instead of turning up the instrument you are concentrating on, look for other instruments that could be brought down or cheated down, especially when driving a bus compressor. My personal mixing practice is to begin mixing without bus compression, only to experiment with it at the end stage of the mixing process — use it only if the glue it provides truly makes the mix sound better.

Bob Olhsson advises:

> I've never heard a compressor or limiter that could beat the sound of manual gain riding in a mix. It's a LOT of work and many people don't have enough time or money, but the results sound huge with just a little limiting on some of the peaks in the master.

We cannot restore quality in mastering that has been lost in mixing. An "open-sounding" mix produces a better-sounding master; punch and impact come from microdynamics. You will still be able to be creative with compression and other effects — *a fixed monitor gain is liberating, not limiting.*

Multichannel Monitor Levels

Multichannel not only means better sound quality with more dimension, it also means more headroom, because six channels produce more sound pressure than two. So the mix engineer doesn't have to push the levels as far to get a forte and the sound can breathe more.

II. A Brief Survey of Monitor Controllers

I have been promoting calibrated monitoring for a long time. That's why I'm pleased to see that several manufacturers have produced monitor controllers that meet the needs of mastering engineers, and one monitor controller which is suitable for budget mix rooms. So audio engineers no longer have an excuse for not having a calibrated monitor controller. All these hardware controllers have built-in DACs which incorporate PLLs with excellent jitter reduction, translating to higher-quality D to A conversion (See Chapter 24).

Hardware Controllers

BMC-2

The economical but powerful TC Electronic *BMC-2* makes a great mix room calibrated monitor control (pictured on page 195). It has nearly all the features mastering engineers need except the ability to automatically offset the monitor gain upon switching sources. However, the clever *reference* button changes the gain to a preselected amount, and with a bit of work it would be possible to switch between the S/PDIF and the Toslink sources and then manually offset the gain, all right for

"A fixed monitor gain is liberating, not limiting."

occasional mastering chores. Since the BMC-2's gain is performed in the digital domain, it's important not to exceed 0 dB gain except for low level sources, and the built in metering would make that clear. It includes a mono/stereo/side monitor. The side monitor helps to reveal abuse and artifacts, especially in overdriven codecs as well as give evidence of the stereo separation of a recording. The BMC-2 can feed either an analog or a digital monitor system and in the latter case not require an additional set of converters.

Crane Song Avocet

The Crane Song *Avocet* (stereo or surround up to 7.1) is a high-quality, sophisticated monitor controller suitable for mastering or a higher-budget mix studio (pictured on page 194). Important features include mono and phase [polarity] switches. The phase switch inverts one channel, and the mono switch combines the channels, which yields L-R (the side signal) when both switches are engaged. To offset the gain for any input, press and hold an input button and then move the monitor control, useful for matching loudness between a mixed source and the master. Many of the other features, including talkback and speaker selection, are more commonly needed in a mix room.

Grace M905

The *Grace M905* (Stereo version) or the *M906* (5.1 version) are high-quality, sophisticated monitor controllers suitable for mastering or a higher-budget mix studio (pictured on page 194). Important features include a mono and L-R switch, and the ability to set up offset gain per input source, useful for matching loudness between a mixed source and the master. A nice fringe benefit is a built in SPL meter. Many of the other features, including talkback and speaker selection, are more commonly needed in a mix room.

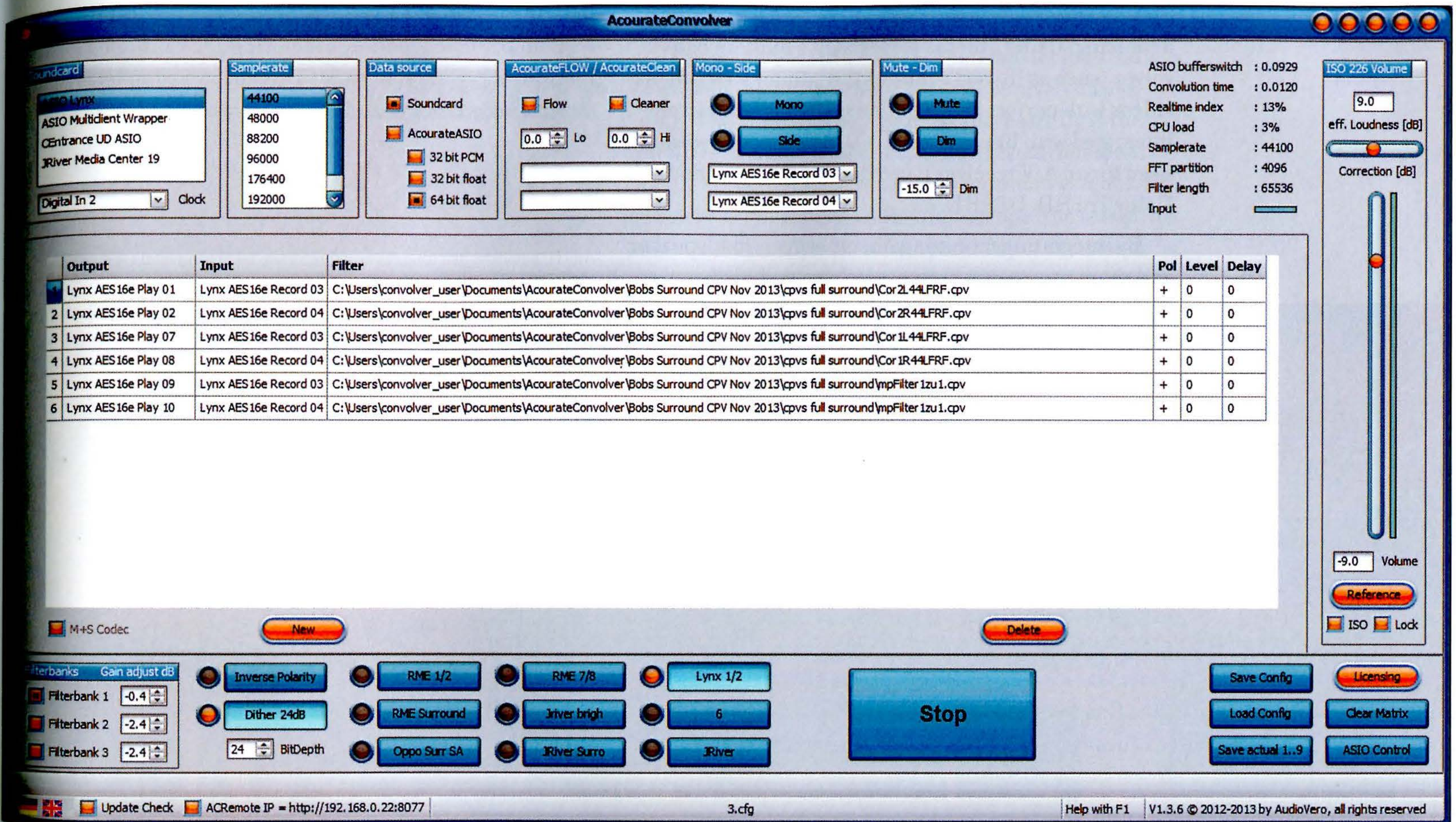
Maselec MTC-1

The MTC-1 includes calibrated monitoring and is also a mastering transfer console, described in detail in Chapter 14.

Software Controllers

In the middle of 2013, I switched over from an analog monitor controller to a very sophisticated digital monitor controller called *Acourate Convolver Pro*, that runs in software on a dedicated computer (pictured on page 271). I made this change with some trepidation: Never turn your back on computers. Imagine having to tell your clients, "excuse me while I reboot my speakers." That's not far-fetched in these days of digital loudspeakers which are potentially a little less stable than their analog counterparts, no matter how hard manufacturers work to make them reliable. Practice the usual precautions of having a clone of the boot drive, or even by building a redundant monitoring computer. I can switch to a redundant monitor control system in case *Acourate Convolver* goes down, but to date that has not been necessary.

Every mastering engineer is concerned about having accurate monitoring, and should rightly question the wisdom of inserting a digital gain control between the source being monitored and the loudspeakers. However, computer power has doubled over 8 times since the previous edition of *Mastering Audio*, and therefore a lot more power, flexibility and signal quality is now available in digital signal processing. In addition, *Acourate Convolver* contains powerful loudspeaker linearization, digital crossover and room correction. I spent months of careful listening comparisons to prove that it sounds as transparent, quiet, pure and accurate as my previous



Acourate Convolver Pro, a software monitor controller with digital loudspeaker and room correction, and more.

controller, but the reason for switching over is that it has actually made my reproduction sound better. I am able to make even more precise judgments about sound quality, dithering, equalization, etc.

However, using a digital monitor controller removes two features: It cannot monitor an analog source without going through an A to D conversion. I listen to LPs and analog tapes through a world-class 192 kHz ADC that sounds extremely transparent. The second loss is the inability to listen to an SACD in DSD mode, but I have replaced that with a high-resolution real-time processor that digitally upsamples DSD to 176.4 kHz/24-bit PCM and sounds very good. An issue is that the convolver runs 65 K-sample FIR filters for crossover and loudspeaker linearization, which have a latency of almost one second at 44.1 kHz. I'm more than willing to live with that latency for the improvement in sound quality. I can switch *Acourate Convolver* to a low latency mode at the push of a button for editing or denoising. Using a media player with a built-in convolver (such as JRiver) removes the latency so it can play videos with perfect lip sync. The system is completely integrated with Blu-Ray and DVD playback and without need for an A/V receiver it can decode Dolby Digital, Dolby Tru HD, DTS HD, etc.

By integrating monitor gain, crossover, loudspeaker and room processing, input switching and configuration, *Acourate Convolver* eliminates many extra stages that would normally have to be added to a system.

Here are some features of *Acourate Convolver Pro*:

- Remote control
- Switch among 9 different PCM sources, which can be multi-channel, at sample rates to 384 kHz, with gain compensation per-source (e.g. between mix and master)
- Automatic clock switching between internal and any external clock source
- No degrading ASRC in the signal path; all filter coefficients are switched when the sample rate switches (See Chapters 22 and 24)
- Linear phase crossover and system phase linearization
- 64-bit floating point internal processing resolution, dithered accurately to 24-bit fixed point on the way to the DACs
- Seamless bass management, crossover and time alignment between stereo subs and main speakers that makes the entire system full range. Frequency response linearization: my system is 3 dB down at 16 Hz, flat + or - 2 dB to 20 kHz with a controlled high frequency rolloff.
- Mono, stereo and side monitoring, dim and mute

In Conclusion

Sound quality, loudness metering and calibrated monitoring go hand in hand! This chapter completes our loudness trilogy. Viva la Revolución!

1
An ITU standard listening room and recommended component tolerances, specified in ITU-R BS 1116-1, promotes monitor quality and assessment of sound quality.

2
Since the loudness meter combines the channels, and is weighted, some slight revisions to the motion picture (theatrical) standard were necessary. The old single-channel narrow-band pink noise signal measured -20 dBFS per channel on a flat RMS meter. This reads about -22.2 LUFS on the weighted loudness meter. The weighting makes this signal read higher on the loudness meter. Keeping in mind that the loudness meter reads one channel 3 dB lower than the combination of both, we must ensure the single-channel pink noise reads -23 LUFS.

Given how much smaller the music mastering room is than the motion picture theatre, it is a good idea to downsize the level a bit to correlate better with our perception of forte in a smaller room. Even so, SMPTE-calibrated motion pictures sound too loud with the monitor control at the 0 dB position.

3
The ATSC pink noise test signal at the links in the ATSC document is dual channel. It reads -19.7 LUFS, so it is very close to the standard I recommend. Be sure to mute one channel at a time or ensure that the panpot you use does not change the gain of each channel. When in doubt, use a loudness meter to measure the digital output of the signal you are sending to the monitor controller.

{ *"Excuse me while I reboot my speakers!"* }

CHAPTER 20

Monitor Quality

I. Philosophy of Accurate Monitoring

The major goal of a professional mastering studio is to make subjective judgments as objectively as possible. The key to doing this most successfully is the intelligent use of an accurate, high resolution monitoring system. A high resolution monitor system is the mastering engineer's audio microscope, the scientific tool which enables the subtle processing decisions required by our art.

What is a High Resolution Monitor?

It seems strange to describe a passive device like a loudspeaker as "high resolution," but not when we consider its acoustic environment. Also, the technical quality of the loudspeaker and power amplifier enable us to judge the inner details of a recording. I apply the term *resolution* to monitoring in the same sense that I used to discuss digital audio in Chapter 15. For example, a high resolution monitor system has few artifacts that would mask the signal we are listening to. It should have low noise and distortion to permit the ear to resolve low-level signal components. The monitor system, which includes the room, must be quiet. It must not distort at high peak sound pressure levels. Time domain interference by the structure of a loudspeaker can be considered a form of interfering noise. Diffraction can be caused by sound bouncing off of the hard edge of a loudspeaker. Chapter 21 will discuss how room acoustics affect monitoring resolution.

Elements of a High Resolution Monitor System

Here are some guidelines for constructing a high-resolution monitor system:

1) Mastering engineers should work with a single high-quality monitor system that they are intimately familiar with. They will then know exactly how its performance will translate to the real world, so they can please the maximum number of listeners. In general, we will not find multiple choice or alternative monitor loudspeakers within a mastering room.

2) The mastering room must be extremely quiet, with all noise-equipment banished to the machine room. No producing ise floor must be better than NC 30,[1] preferably NC 20 or less.

3) There should be no significant obstacles between the monitors and the listener within the standard equilateral monitoring triangle. If possible, the listener should be within a *Reflection Free Zone* (RFZ), meaning that reflections from nearby surfaces arrive at the listener at least 20 ms later than the direct sound, and be at least 15 dB down. In Chapter 21, I will document RFZ (RFZ coined by Dr. Peter D'Antonio.)

4) The electronic chain should be designed for maximum transparency. Often we use specialized or customized analog components that incorporate a bare minimum of active stages.

5) Monitor loudspeakers must have low frequency response flat to 20 Hz or below, wide bandwidth, high headroom, and extremely flat frequency response (up to about 2 kHz, where they begin a slow measurable rolloff). An inaccurate monitor system can lead to inappropriate EQ adjustments. An accurate monitor system reveals differences between recordings: A system that makes everything sound the same, or makes everything sound beautiful, cannot be accurate.

6) Monitor distortion should be very low to avoid monitor compression. In Chapter 21 we'll present benchmark measurements of actual monitors.

7) Time-domain problems, including sources of diffraction must be minimized. Cabinets should be solid, non-resonant, and free of sympathetic vibrations and resonances. Sweep a sine wave at 80-90 dB SPL from 20 Hz to 250 Hz to test.

8) The walls in the room must be solid and non-resonant; the room large enough to permit even, extended, bass response, with no significant standing waves. Any remaining standing waves should be controlled using traps or specialized diffusers. Ideally, room length should be at least 18 feet long (5.5 meters), preferably 30 feet (9.1 meters). The shape of the room should be symmetrical from left to right. The ratios of length, width and height should not be integer-related and carefully calculated to avoid accumulation of standing-wave frequencies (see links for an online calculator). All loudspeakers reproducing above approximately 100 Hz should be equidistant from the listener, with possibly digital delay distance compensation in the surround speakers if the room geometry does not permit equidistance. The room should be wide enough so that first reflections from the side walls arrive at the listener's ears significantly delayed and attenuated. The side walls, ceiling and floor should be treated to minimize **specular** reflections arriving at the listener's ears. The rear wall should be far enough back and/or treated to minimize specular reflections' amplitude and maximize the delay before its reflections reach the listener. All objects in the room should undergo similar consideration.

9) Acoustical design and electrical layout should be performed by experienced and trained professionals.

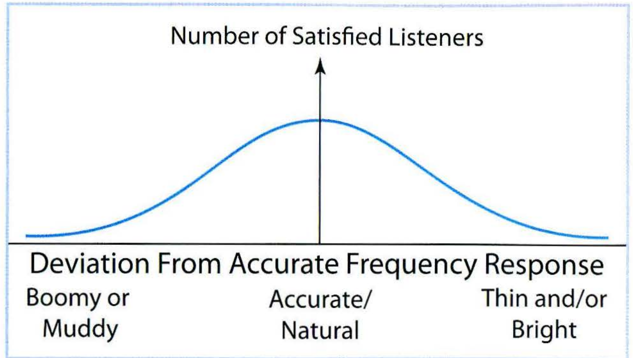
Subwoofers and Bass Response

Stereo subwoofers, or prime loudspeakers whose response extends to 20 Hz are essential for a good mastering studio. Vocal pops, subway rumble, microphone vibrations, distortions and the lowest notes of the bass will be missed without extended low frequency response. Subtle judgments of bass drums can be better made using a loudspeaker with infrasonic response. Sometimes I introduce a high pass filter at 10 Hz, a judgment that can only be appreciated in a few listening environments, but at least the judgments that I can make are not crippled. Proper subwoofer setup requires knowledge and specialized test equipment (see Chapter 21). If subwoofers are inaccurately adjusted (e.g., "too hot", in a misguided attempt to impress the client) then the results won't translate well to other systems.

Apparent bass response is also greatly affected by monitor level. Don't listen too softly or too loudly. The equal loudness contours (originally studied by Fletcher and Munson, updated in ISO 226) dictate that a recording that is mastered at too high a monitor level will seem bass-light when auditioned at a lower level in a typical home environment, and vice versa. I believe a good monitor level in a professional mastering room that will translate is about 83-86 dB on forte passages.

Why Accurate Monitors Are Needed

Because the mastering engineer strives to create a recording that will play well on the maximum number of playback systems, it is best to work to the middle of a bell curve, as illustrated in the figure (at right). It is obvious that tilting a recording in the bright direction means it will not play well on a lot of small systems that already have too much treble, or that skewing it in the direction of the bass means it will not play well on systems that have too much bass. The more linear the response in the midrange, the better the sound will translate, because clearly there will be systems with some anomalies. But most speaker designers seek neutrality.



An accurate monitor system allows us to produce recordings which are in the middle of the curve.

II. Debunking Monitor myths

Myth #1: You must mix (master) with real-world monitors to make a recording for the real world.

Some problems I have heard in mixes can be directly traced to monitor coloration that is endemic in some "real-world" speakers:

- The bass drum is boomy (probably caused by console resonances, dips, and peaks in the near field environment)
- The vocal is too low. Caused by center buildup in the near field environment, or by engineers concentrating on the sound of the instruments instead of the holistic balance.
- The reverb is too loud (probably caused by low-resolution monitoring or a noisy listening environment)
- The bass instrument level is too low compared to the bass drum. Usually caused by depressed mid-bass due to comb filtering artifacts from a console reflection, or by using this month's flavor of mix loudspeaker

"It's time to stop catering to loudspeaker fashion and move to accuracy."
— Bob Katz

MYTH: *Program Compression is required to protect small reproduction systems.*

- The presence or high end is greatly exaggerated (caused by the low-resolution monitoring environment obscuring inner details and tempting the mix engineer to boost an equalizer)
- The stereo separation is very small (nearfields exaggerate separation like a big pair of headphones)
- The high end is, at best, unpredictable

Using the current loudspeaker flavor of the month influences the mix engineer's decisions. But next month some other non-linear loudspeaker will be touted, and it will be time to change the fashion. So, which "real world" loudspeaker should the mix engineer choose? It's time to stop catering to *fashion* and move to *accuracy*. Mastering engineer Glenn Meadows, who is also an advocate of getting mixing engineers to use accurate monitoring, sums up the situation very well:

> The biggest trouble is when engineers second-guess what they're hearing and try to correct for monitoring/room issues. Then there is no anchor...
>
> As a mastering engineer, maybe most of my job is to dial in a correction curve to bring out what the engineer heard.

I see mix engineers struggling every day switching between monitors, trying to guess which one is telling them the truth, or being fooled by the one that tells them the most information at any given time. In our mix room, we have one pair of accurate monitors, and can find the right mix faster and easier than an engineer with multiple monitors. The mix translates everywhere with very little effort. This works!

Myth #2: Adding high end helps a recording translate to home systems.

This is an untruth, because adding high end (more than what sounds right) skews a recording away from the mean of the bell curve, resulting in a sound that's sharp, thin, tinny, and fatiguing. Its vocal/instrumental balance will be wrong on ordinary speakers in ordinary rooms. Radio play will suffer, because FM radio limiters will cut back the added highs. But most important, if the midrange is wrong, nothing else is likely to be right.

Myth #3: Heavy compression is necessary to prevent small monitor systems from overloading.

I have found the opposite to be true. When I take my dynamic masters to a consumer type Aiwa 3-piece system, they sound (comparatively) compressed, with fewer transients and less impact. If I had reduced the transient clarity in the mastering, it would only sound worse on the smaller system, which performs its own compressing! Thus I have learned that even if the sound "sticks out a little too much" on a high-headroom mastering system, it's probably going to be fine when played on an inferior system. And vice versa, if we make judgments on a compromised monitor system, we'll never learn if something is over- or under-compressed. This is another reason why mastering engineers reject typical near-field monitors, because very few available near-field monitors pass the bandwidth and compression test, or can tolerate the instantaneous transients and power levels of music without monitor compression. Even so, placing monitors in the near field exaggerates transients by greatly reducing the contributions of the room, in a way that does not happen in the consumer's listening world.

III. Refinements

Alternate Monitoring Systems

Mastering engineers use alternate loudspeakers as a double-check, not as a benchmark. With the proliferation of portable media players, headphones make a good double-check, although they tend to exaggerate inner details of a mix. We have another room whose system has large, "loose-sounding" woofers, representing an extreme of the bell curve (which may happen to the bottom end in a club, or car). Cars are likely to have very uneven bass response, aggravated by user-set equalizers (see photo at right).

I've learned to watch out for recordings where the client is looking for very hot bass or bass drum. If we boost the bass on the master in the neutral mastering room to get the distorted bass sound they're used to, it will overdrive a typical car system. The boomy alternate listening room demonstrates to them what can happen. I try to recommend that the customer not use the car as his reference, because it invites him to turn up the treble, due to the limitations of the car, and if they make a recording sparkle in the car, it will screech at home and overload the FM radio processors when broadcast.

Narrowcasting

There are boombox, club, and car systems whose bass response/resonance is so extreme, they should not be included in the bell curve. It is very difficult to make a single master that plays well on a club system with exaggerated 24" woofers that doesn't sound thin and lifeless on all other systems. Making a master that translates to all environments is an art, especially if the club is included. Sometimes it's necessary to make a separate (dedicated) master for club playback, if the customer wants to have a big, loud, boomy low bass. I've also found more than one client auditioning on MacBook speakers ask for more bass, a sure path to disaster!

Here's what we're up against

Monitor Equalization — by Ear or by Machine?

Initially I was against the use of external monitor equalizers to linearize a loudspeaker system, because I found that they reduced the transparency and introduced more problems than the loudspeaker's analog crossovers and drivers. I previously concluded that it was not possible to introduce digital correction in a chain that would not obscure our judgment of fine details in a recording. Since the first edition of this book was released, all of that has changed. Moore's law has seen eight doublings (one every two years) in the number of transistors on a chip — and my approach has turned around 180 degrees!

"The car system should be a double-check, not a benchmark."

"A monitor that makes everything sound beautiful cannot be accurate."

As technology has improved, digitally-corrected loudspeaker systems of high quality suitable for mastering have begun to appear, such as the PMC model twotwo•8. The key is to begin with loudspeakers engineered with high-quality components that sound and measure well as a system, before any digital correction: then to apply the digital correction as a "polish." Linear phase digital crossovers with very high resolution can now be constructed that are sonically superior to previous analog crossovers.

After over a year of experimenting and shootouts, I have finally converted my mastering monitor system to a digitally-corrected loudspeaker system. Believe me, I did not take this radical step lightly. In the years to come, I am sure more mastering engineers will come around to this approach, but in the meantime, the vast majority of conservative mastering engineers will stick to an all-analog, uncorrected monitor system.

The gain structure of 24-bit DACs and monitor amplifiers can be adjusted so there is plenty of digital headroom for peaks and still an inaudible noise floor, even with a digital monitor level control. When properly dithered its noise floor acts exactly like an analog system. That is, as signals get quieter, they gradually disappear into the noise without any quantization distortion. The digital crossover is in fact quieter and more transparent than previous analog active crossovers; its signal-to-noise ratio is superior to that of an analog-domain active crossover. It has less distortion than passive or active analog crossovers with their physical inductors and capacitors. The levels in each crossover band (e.g. low-pass, high-pass) will never peak overload, because each band covers only part of the signal, and so has even more headroom than the full-range digital signal.

We know how to measure and equalize for a high-frequency rolloff using a variable FFT window. We can predict a suitable rolloff within a couple of dB, to allow for variance in room acoustics and loudspeaker polar response (try starting at -6 dB at 20 kHz). I use an empirical method of judging the high-frequency response by listening to about 50 of the best recordings that I know. Previously I tried and abandoned using an analog equalizer to fine-tune this HF response, because it reduced the monitor transparency. But now, with transparent, high resolution digital equalization, I have arrived at a very smooth high-frequency rolloff that meets my desired target. Digital correction adjusts the high-frequency response exactly where I want it. I refined this response to my listening goal by using these 50 best recordings. The results are superior to any analog correction system I have previously tried. We'll tell the story of this digital correction system in Chapter 21.

IV. In Conclusion

Even the best master will sound different on different systems, but it will sound most correct when it is approved on an accurate monitor system. That leads us to this comment from a good client:

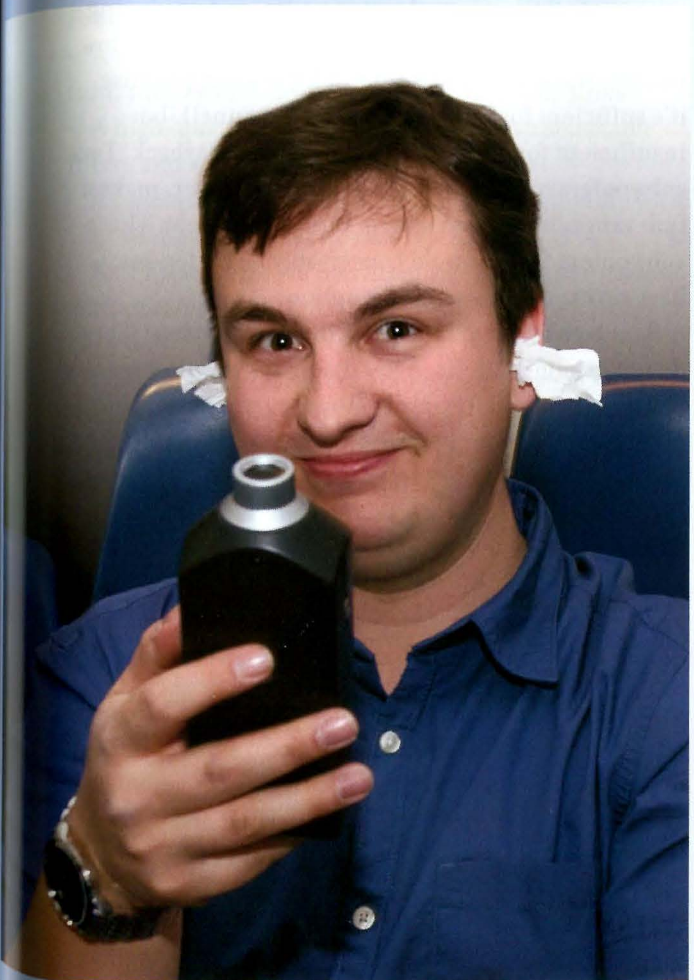
> I listened to the master on half a dozen systems and took copious notes. All the notes cancelled out, so the master must be just right!

---

1 NC 30. Noise criterion 30 decibels, follows an attenuation curve whereby at 2 kHz, noise level is 30 dB SPL, and at lower frequencies is permitted to rise.

CHAPTER 21

Monitor Setup



I. Introduction

In Chapters 19 and 20 we learned why calibrated monitoring and monitor accuracy are so important to the mastering engineer. In this Chapter we will look at objective tools and methods for calibrating and verifying our monitors. Most mastering engineers use the traditional approach of picking superior full-range loudspeakers with their built-in analog-domain crossovers, placing them in well-designed rooms, then optimizing the loudspeaker placement by a combination of ear, calculation, and measurement. Every room has bass-region compromises, though the larger well-designed rooms have fewer of them. My room (like most) has bass anomalies caused by modes, e.g. between the front and back walls. In this room, locating the loudspeaker is critical: the wrong location improves one frequency area, but degrades another. I determined long ago that in my room it is better to separate the woofers physically from the mains, to allow for optimizing the woofer position while preserving the depth and soundstage provided by the main speakers.

After years of incremental improvement, the analog-domain crossovers and low frequency EQ I had built to linearize the response in this room reached their limits: the steeper the crossovers, the flatter the frequency response, but the worse the phase response, causing issues of time alignment between the woofers and the mains and creating a less coherent system. It was clear that positioning the woofers in the plane of the main speakers was a compromise. In my room, measurements showed the best woofer location was the corners, but that would cause a time delay between the woofers and the mains, unless we employed the radical step of digital-domain time correction. I use the word "radical" only because time-domain correction is not a commonly-employed tool (yet), but keep in mind that analog-domain crossovers greater than first order have phase shift, time smear and irregularities in frequency response.

Digital-time correction requires separate DACs for the mains and the woofers, and if possible a digital crossover — elements that are becoming common in a growing number of respected "digital" loudspeakers. However, "rolling your own" crossover is not common. First I sought out a solution that meets my high standards for transparency, resolution, musicality, depth of image, and purity of tone. Most approaches I encountered lowered the sound quality in some way, either by having poor resolution or overcorrecting. Digital monitor equalizers must be of the highest quality. This is why I mention only specific brands and techniques in this chapter that have worked for me. Like all tools, digital correction must be used properly. This chapter is not a detailed lesson in how to create a digitally-corrected loudspeaker, but it does present examples of how it can be implemented and measured.

II. Summary of Essential Tools

- An acoustical consultant or acoustic architect to help plan the room and/or the placement and choice of the loudspeakers.
- A great room, with proper dimensions, wall, floor and ceiling construction, layout, and interior treatment. There should be minimal obstructions/reflections between the loudspeakers and the listener, with low noise and good isolation from the outside world.
- For 5.1 surround sound, five matched full range, "satellite" or "main" (as I call them) loudspeakers and amplifiers with flat frequency response (preferably good to 60 Hz or below). High headroom monitors are necessary to make proper sound judgments: if our monitors are compressing, we cannot judge how much compression to use in the recording. Each main loudspeaker should be capable of producing at least 100 dB SPL at one meter distance with less than 1% total harmonic distortion above 100 Hz. Distortion should be less at lower levels.[7]
- For mastering, if subwoofers are needed, two subwoofers, capable of extending the low frequency response of all the mains down to 20 Hz, and producing at least 100 dB SPL with less than 3% distortion below 100 Hz. Having only one subwoofer is a compromise: it's sufficient for the LFE ("point 1" channel), but insufficient for great stereo or surround playback. Two subwoofers, properly placed and crossed over, make a full-range, integrated stereo/surround system. Three subwoofers (LCR) would conceivably be better, but I have no experience with this option!
- Bass management, needed if the main loudspeakers do not extend to 20 Hz. A low distortion monitor matrix with versatile and flexible bass management, capable of repeatable, calibrated monitor gains, down-mixing, and comparing sources from 7.2 through mono.
- A monitor selector to feed the matrix.
- Measurement/calibration equipment.

For initial alignment of the room, the most critical ingredient is knowledge. A trained acoustician or acoustical architect should help with the first-time setup. He will examine the dimensions and construction of the room and recommend loudspeaker placement and trapping, absorption or diffusion, if necessary. Once the room has been set up, he will perform near-anechoic and early-reflection analysis, adjust subwoofer levels and crossovers, and help solve room response errors. If the speakers are free-standing (as opposed to **soffit**-mounted), he will help find an optimum position that produces the flattest frequency response and best stereo imaging. Hire a consultant but be a knowledgeable consumer: I suggest furthering your acoustic education through the links at digido.com. For example, visit Hunecke in the links for an excellent online loudspeaker position calculator to help position loudspeaker and listener for the flattest frequency response, and a room modes calculator.

Tools for Level and Frequency Response Calibration

- A narrow-band pink noise signal for level calibration, such as the one downloadable from the links
- A high-quality calibrated *SPL meter* with selectable filters and response speed
- A *calibrated measurement microphone*, preferably with 1/4" (6 mm) diameter or smaller capsule
- An FFT analyzer such as *Acourate*, and a filter such as *Acourate Convolver*

or

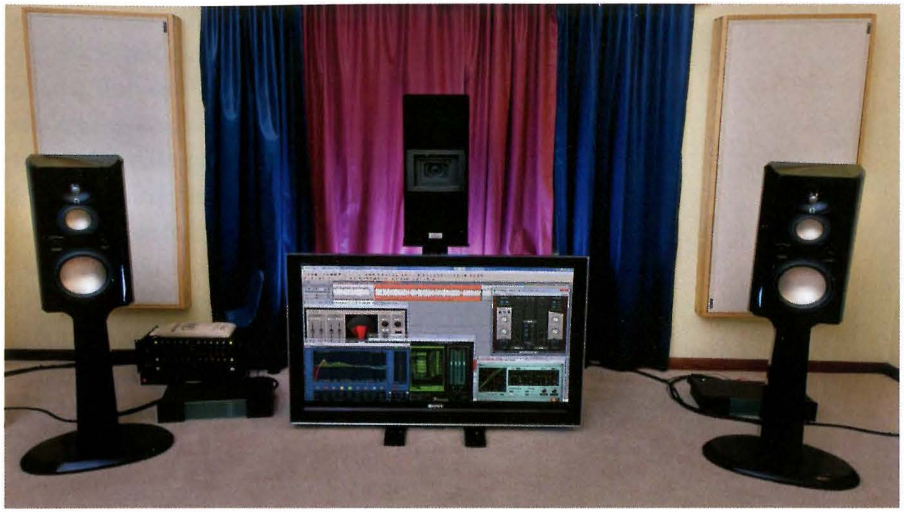
- The excellent shareware application *Room EQ Wizard*, coupled with other equalization or crossover solutions. The help menus and tutorials in REW are superb introductions to practical acoustics and the use of an FFT analyzer.

III. Acoustical Considerations

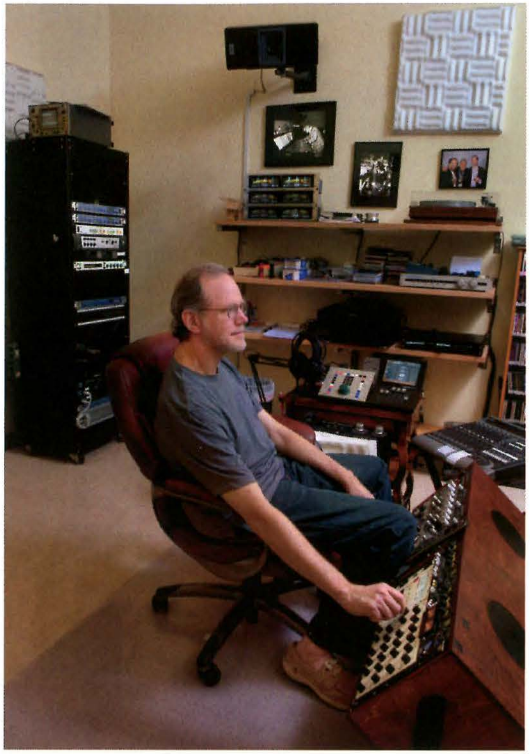
Controlling Obstacles and Eliminating Reflections

The ideal reproduction system should have no obstacles in the path between the loudspeakers and our ears. A trained acoustician can test for the effect of obstacles in the studio environment. Reflections within the first 15 to 20 milliseconds after the direct sound hits the ears should be below -15 dB, preferably below -20 dB. This is called a *reflection-free-zone (RFZ)*.

One of the important features in Studio A is the wide room, 14 feet (4.2 meters). This reduces the significance of direct (specular) reflections from the tweeter for several reasons: the increased distance decreases the amplitude, the angle comes from the side of the tweeter which has lower output, and the delay is greater, taking the reflection away from the early zone where comb filtering could become a problem. Move a mirror around on the side wall until you see the reflection of the tweeter while you're sitting in the sweet spot. At that location, put some treatment. Your acoustician will help decide whether diffusion, absorption, or some combination is best. Another feature in Studio A is a cathedral ceiling that begins at 12 feet (3.6 meters) in the front and angles up to 23 feet high (7 meters) in the back, which virtually eliminates low frequency modes and specular reflections, since the ceiling and floor are not parallel. Another feature (pictured above) is a custom video monitor mount that places it directly on the floor between the speakers, well below the major drivers. We've performed measurements and listening tests with and without this video monitor — it is sonically undetectable and its presence virtually unmeasurable.



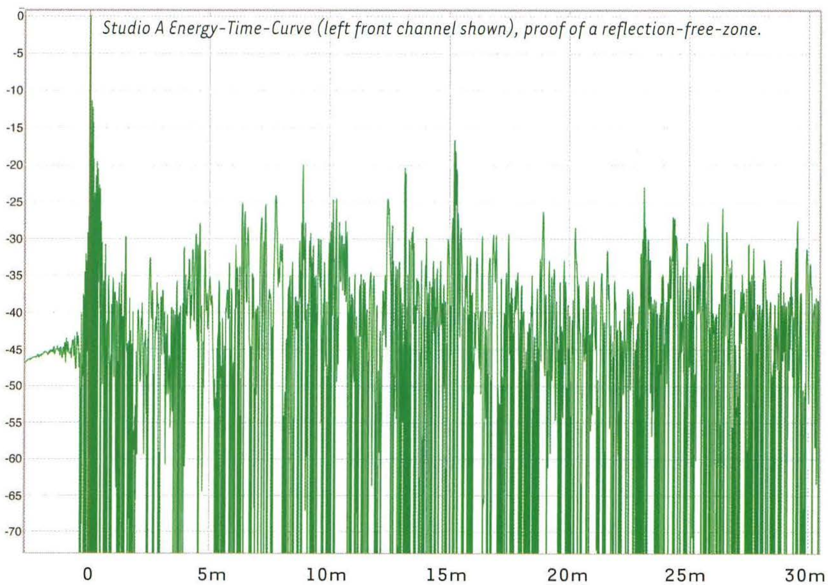
Custom video monitor mount places it on the floor, avoiding interference with the sound produced by the main speakers.

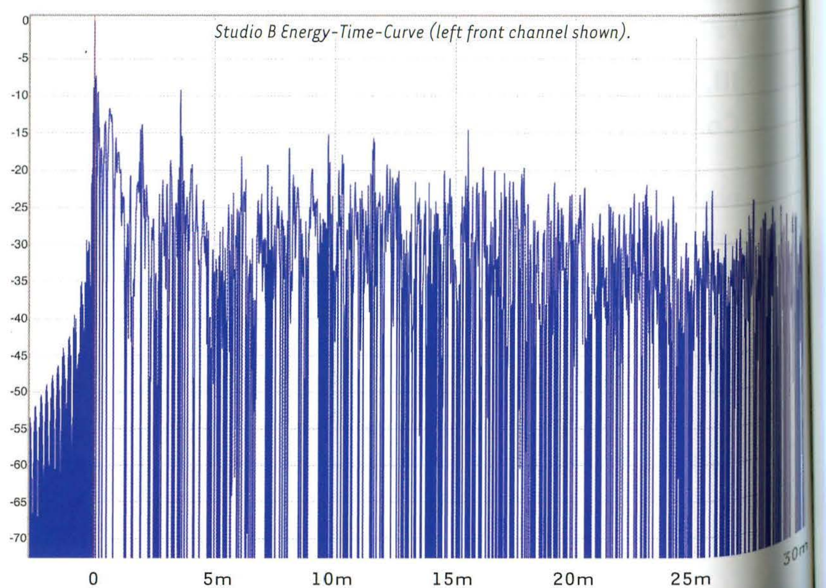


Reflection-free zone

Here is Bob sitting in the reflection-free zone. Notice that the custom-made "low-boy" rack avoids obstructions and reflections between the loudspeakers and the ears. The remaining nearby objects are also carefully placed to avoid a specular reflection at the ears, but the proof is in the **ETC** (Energy-Time-Curve) measurement, pictured below.

The initial impulse from the loudspeaker is set to 0 dB and 0 ms for reference. Vertical scale is 5 dB per division, and time on the horizontal axis is marked in 5 ms increments. In the Studio A measurement (green trace), notice how quickly the sound decays after the initial impulse; in less than a millisecond it has already arrived at -20 dB, it arrives at -35 dB before 2 ms, and it remains below about -30 dB for almost 5 ms. This very sharp impulse is a result of good acoustics and digital loudspeaker correction. Then the room sound remains below -20 for many ms thereafter, which creates a distinct gap before the onset of the dense room modes, eliminating comb filtering and helping to clarify the direct sound. For 30 ms, the sound level does not rise above -17 dB. Except for one somewhat diffused reflection around 16 ms, it would be below -20 dB and largely below -25 dB, which is spectacular performance. A reflection at 16 ms means that the sound has bounced off a surface about 8 feet (2.4 meters) past the ear and then returned. This is the diffused bounce off the back wall, which is sufficiently delayed and attenuated. This bounce clarifies the direct sound and "decodes" the ambience and depth in a recording due to the Haas and Madsen effects (See Chapter 10). You should aim for this quality of ETC in a mastering room or any room seeking ultimate sound definition.





Reflections are serious business. Poor time-domain response means poor frequency response. The sound of Darth Vader's voice in Star Wars is a comb filter created by mixing in a delay at 10 ms with the dry sound.[1]

The second ETC (blue trace) is Studio B, our mix room. This curve looks like the impulse from a conventional loudspeaker in a small room with a console surface. Notice that the initial impulse is spread: the sound in Studio B takes a much longer time to decay to -20. Interfering reflections are immediately present within the first 5 critical milliseconds. Reflections through the first 30 ms are much denser and louder than in Studio A, characteristic of a much smaller room in which the sound can never be as clear. Although the console in Studio B is carefully located and angled to avoid pointing reflections to the ear, it still causes measurable and some audible interference. Nevertheless, this is better than many ETCs I've seen for a small mix room with a console. Studio B's racks are built low to avoid reflections to the ears, and the video monitors are mounted flush with the console table, avoiding the common mistake of placing video monitors in the way of the audio image. A poor ETC translates to poor frequency response; evidence of comb filtering in the lower midrange can often be seen in such rooms. So if you don't want Darth Vader in your monitors, heed the evidence of the ETC!

Room Treatment

Before thinking of employing any correction or equalization, first do everything possible to fix room-induced problems by using acoustical treatment and properly locating the loudspeakers. Small-room acoustics is divided into low frequency problems and high frequency problems at the so-called **Schroeder Frequency point**, below which the room behaves largely modally (room modes are standing waves at particular wavelengths that are integer-related to the distance between walls). Analyzing low frequency problems includes taking waterfall measurements to examine room modes (See links). Trapping (and sometimes low frequency diffusion) is required to deal with low-frequency room modes. An untrapped room will exhibit resonances, peaks and dips at modal frequencies, which make the sound boomy, blur the bass and cause a loss of definition in the sound. But trapping is often misapplied, because low-frequency wavelengths are much larger than the dimensions of most practical traps. It takes more than a few inches of absorbent material to dampen a resonance at, say, 50 Hz, and those few inches can easily overdamp the mids and highs. The result of improper or excessive trapping is an overdamped room, where the sound has lost its liveliness. To keep the sound lively and natural, ensure that absorption is not excessive in any octave band: sound should not decay significantly faster at high frequencies than at low, or the room will sound dead. The measurement of octave band decays is called the *Schroeder Curve*. I've been experimenting with the Bag End E-Trap which, unlike typical absorbers, does not dampen frequencies we want to leave alone. The E-Trap helps keep the room lively.

Why Stereo Subwoofers?

Stereo subwoofers avoid the compromise of a mono subwoofer for several reasons:

- Stereo subwoofers provide a greater sense of envelopment than a single woofer, even when reproducing mono material. There is also evidence that subwoofers can be localized to some degree.[3]
- Stereo subwoofers help create the same effect as full-range stereo loudspeakers.

• Stereo or multiple mono subwoofers can average out and help to reduce modal buildup in the room.
• Stereo subwoofers double the headroom, 3 dB more power than a single subwoofer of the same power.
• Frequency response is always a compromise with a mono sub: it will never be correct for all sources. This is because low-frequency levels are different when combined electrically as opposed to acoustically (in the room). For example, two channels of a center-located in-phase bass instrument combined into a single (mono) subwoofer results in a 6 dB increase, whereas with two separate subwoofers, the sum is between 3 and 6 dB depending on room acoustics and the distance between the speakers. If only a single (mono) subwoofer is used, center-located bass information will likely sound a little loud. So, if stereo subs are used, then be sure to check sound in mono when mastering, to hear what the effect on the bass response would be in a consumer system using only a single subwoofer.

Alignment: Lasers and Time Delays

Our goal is to adjust the loudspeaker angles and precisely match the distances to conform with the ITU 775 recommendation[4] (pictured on page 285), or a variance that we may prefer (see Chapter 11).

We start with calculated marks on the floor for the left and right loudspeakers and listener, but we may need to fine-tune their precise locations. With the room interior acoustics reasonably treated, place the loudspeakers and listener at their proposed locations and do a preliminary listen to some stereo material. If you hear a nice image and depth, then this is indeed a good starting point. Confirm the actual listening location by dropping a plumbline[2] from the tip of our nose to the floor, make a mark, measure back 5 or 6 inches (about 14 cm) for the location of the ears, and if necessary revise the position of the provisional mark. We reconfirm that this mark is exactly centered between the two side walls, then put a marker on the floor or carpet at this spot. All speaker angles and distances will be precisely measured from this spot.

The ITU 5.1 setup (pronounced "five point one", five main channels plus LFE) is a circle, with each loudspeaker at identical distance from the listener. ITU's recommended surround speaker angle is a compromise between good localization and good ambience. The ITU recommends between 100-120° for the surrounds. Rest the indent of a laser "chalk" at the mark on the floor. It helps to put a temporary nail in the floor at this position to precisely align the laser. Shine the laser line along the floor to the front wall. Rotate the laser line until it is precisely perpendicular to the front wall, then put a narrow piece of tape on the front wall at that center spot to sight the laser line in the future. Now place the center of a surround speaker alignment chart[5] over this spot and place the laser chalk over it, as in the photo (page 285). With the laser line pointing at the center tape on the front wall, adjust the chart until it's centered, and tape it down so it doesn't rotate.

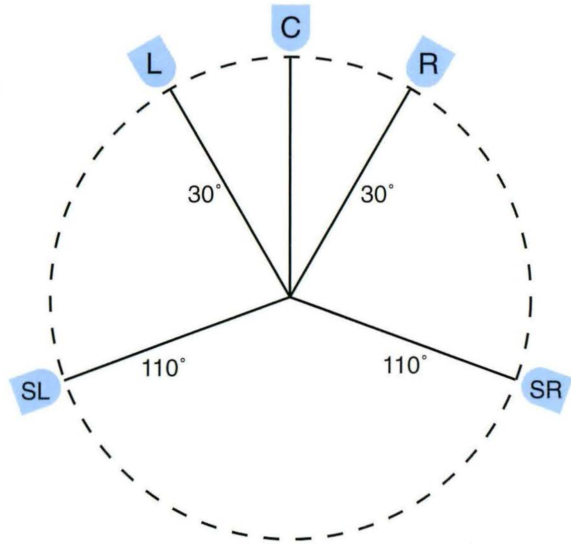
Now point the laser line at each of the chart's speaker positions and place the acoustical center of the loudspeaker at that angle, and at the distance proposed by the acoustician for the flattest response in the room; if it helps, drop a plumbline from the speaker so it lands on the laser line on the floor. It helps to "toe in" each loudspeaker so the front of the speaker is perpendicular to the laser line. Later, this angle can be tweaked during listening; the toe-in helps off-center listeners hear the opposite channel, but too much toe-in can defocus the phantom center for center-located listeners and skew the shape of the soundstage image.

The next tool is an electronic distance calculator, the delay measure of a time-domain (FFT) analyzer. If it takes identical time for sound to travel from each speaker to the microphone, the speakers must be at identical distances. Adjust all the distances precisely until they match electrically. If it's not possible to put the surround speakers at the same distance from the listener as the fronts, then it may be necessary to insert a time delay on the appropriate speaker, which will be handled later by the correction system.

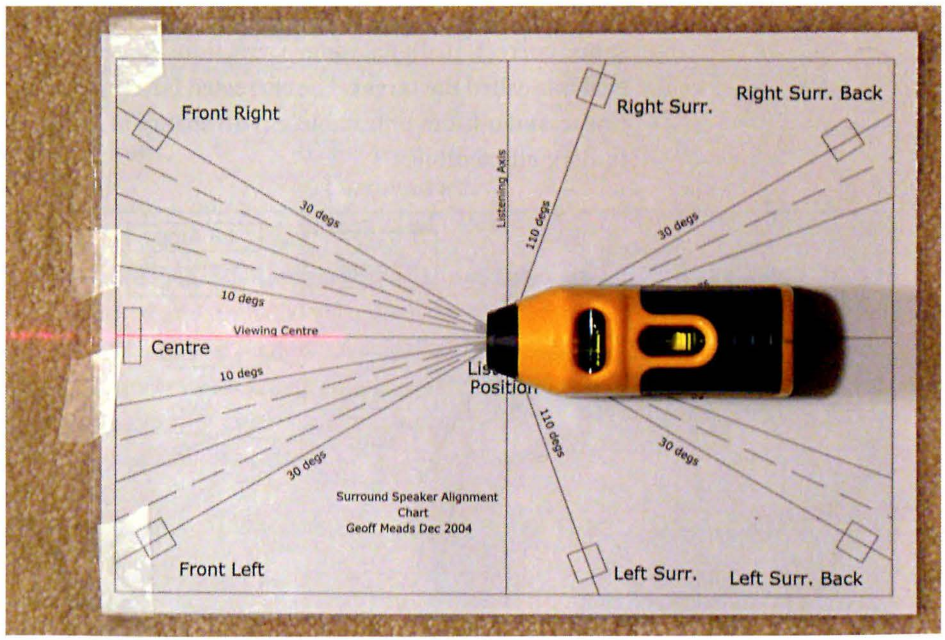
Place the subwoofers where they have the flattest measured response if you are going to implement a digital time delay for them. Using an FFT analyzer, move one subwoofer around until the measured differences in peaks and dips between about 50 and 100 Hz are minimized. If you are not going to implement digital time delay, measure the time arrival of each sub and try to place them so it matches the main speakers. Aligning the subs physically with the mains helps to produce a coherent image, but not necessarily the flattest low frequency response. It's much better to place them in the best location for frequency response, then compensate with a digital time delay. Pictured (page 286) is a JL Fathom F112 subwoofer, which we use to combine the LFE with bass management.

IV. Taking it Beyond: Digital Monitor Correction and Crossover

One clue that a loudspeaker system must be inaccurate is that during mastering we are continually tempted to fix problems in the same frequency range. This means that the problem lies with the reproduction system, not the recording. The process of analyzing and correcting a system requires a skilled acoustician who understands the tradeoffs of electrically equalizing the direct response — for example, when a room anomaly is the root cause. Although equalization can be totally correct only for one spot in the room, waterfall plots reveal that careful low-frequency equalization can be very effective at reducing some room-caused anomalies. Visit the discussion on this topic in the links. For high frequencies, I have always relied on the natural rolloff of a well-made loudspeaker to make the 50 best recordings I know sound as correct as possible. Now I use digital response correction to



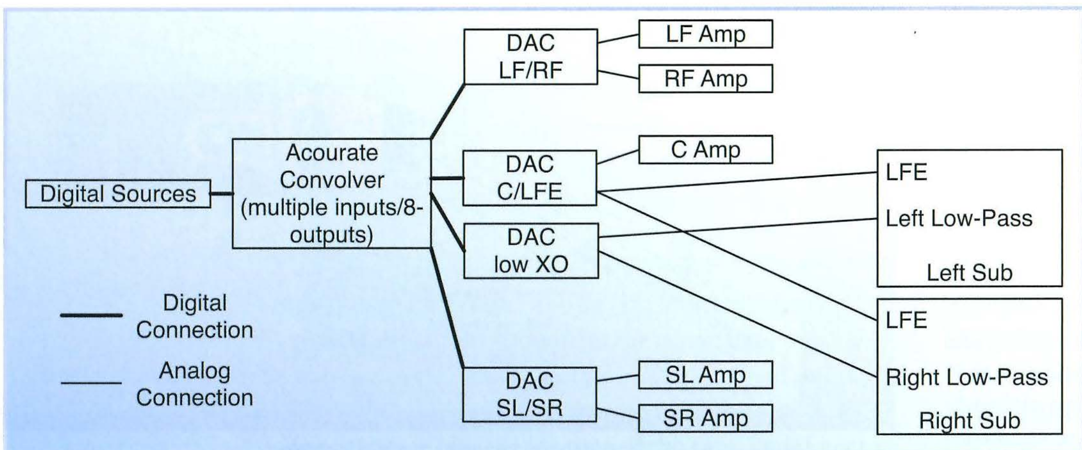
The ITU-R BS.775-1 recommendation for 5.1. Tolerance on surround angles is between 100-120°.



Laser pointer resting on an alignment chart

JL Audio Fathom F112 Subwoofer

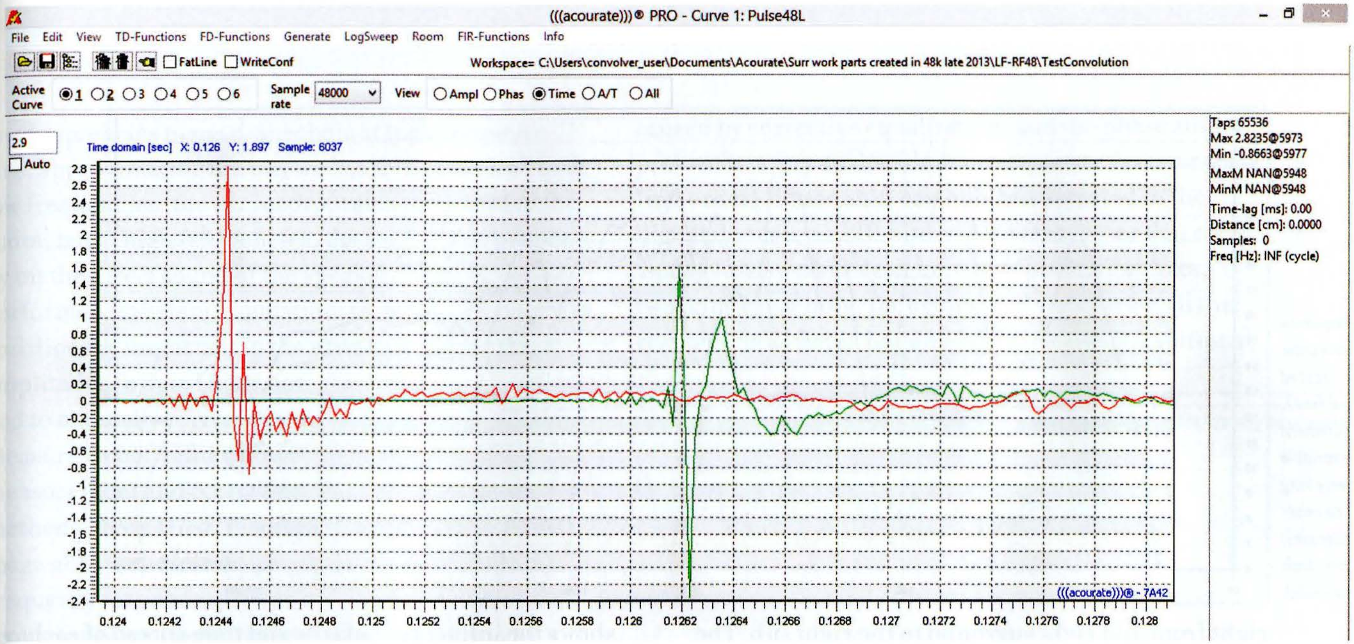
refine that high frequency rolloff till these recordings sound correct. In digital room correction, the response goal is is called the **target**. The corrected HF response is linear, smooth and repeatable: 0.1 dB change at 20 kHz is distinctly audible.



Digitally-corrected 5.1 monitor system with a 5x3 digital 2-way crossover

The monitoring system is the mastering engineer's audio "microscope." Acourate introduces transparent digital components that remove errors (such as frequency response anomalies and phase shift) caused by analog components (including drivers and crossovers) in the monitoring system. It becomes a better microscope without losing any of its transparency, gaining improved accuracy and purity of tone, impact, transient response, imaging, and depth.

My target is flat until 1 kHz (the hinge point), followed by a linear rolloff to about -6 dB at 20 kHz. Response is measured using a psychoacoustic measurement method that includes a sliding FFT window, and a proprietary technique that avoids overcorrection (too much equalization) by performing thousands of transient simulations to find only the frequencies that truly need correction. Your choice of high-frequency rolloff may differ, since the amount of rolloff depends on the room treatment, the polar pattern of the loudspeakers, and your listening preference. But everyone prefers some amount of measured high-frequency rolloff. After smoothing out loudspeaker anomalies, you will have a system that reveals any problems in recordings, permitting professional judgments. A linearized system is both revealing and good-sounding at the same time: Because non-linearities in the loudspeakers have been removed, a larger portion of the recordings in a collection sound good. There is less chance that a resonance in a music recording would coincide with an unwanted resonance in the loudspeaker or room.



Uncorrected impulse response of the front left speaker (green). Corrected impulse (red).

Connecting and Calibrating the System Levels

Pictured (page 286) is a block diagram of a digitally-corrected "5 point 1" monitor system with a 5x3 digital 2-way crossover. There are 5 high pass outputs and 3 low pass outputs, hence the designation "5x3." Multiple digital sources enter the inputs of a computer interface controlled by *Acourate Convolver Pro* (See Chapter 19). After the digital crossover and DACs, multiple analog outputs feed the respective loudspeaker amplifiers. Notice that the LFE (a single channel) is split with a Y-cord to feed both subs.

Acourate Convolver Pro is

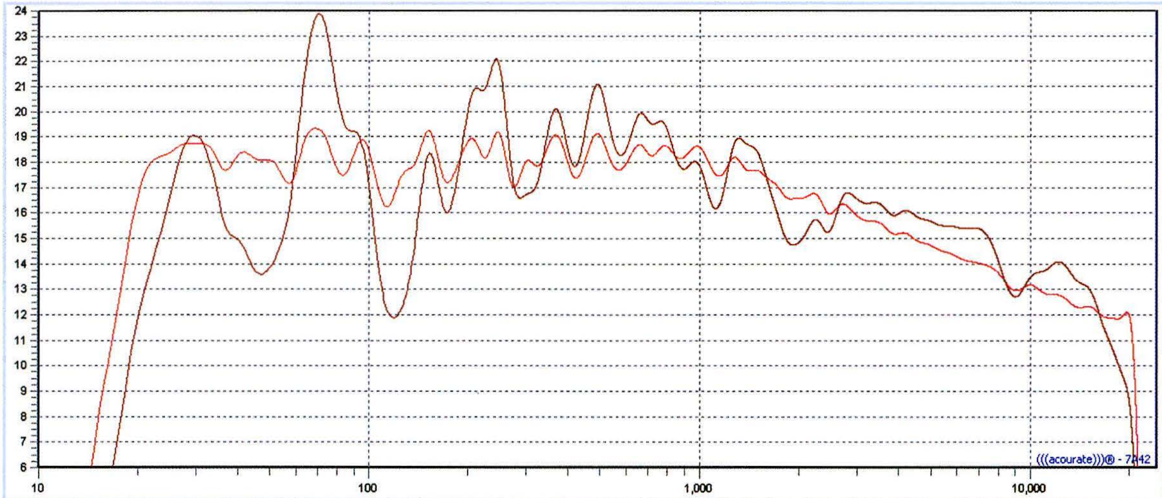
- an input selector for different digital sources
- a 64-bit digital filter, optimizing the impulse, phase and frequency response of the system
- a digital crossover (bass manager), dividing high pass and low pass between main speakers and the stereo subs and time aligning all loudspeakers
- a monitor level control, dithered to 24-bits to properly feed a DAC, with dim, mute, sum and difference check

My studio is configured for 8 digital output channels, which enter four stereo DACs:

1) Left front	5) Left low pass
2) Right front	6) Right low pass
3) Center	7) Surround left
4) LFE	8) Surround right

The built-in analog crossovers in the subwoofers are bypassed and replaced with the linear phase digital crossover. Each subwoofer can mix multiple analog inputs: in this case we use two inputs of each sub, one for the bass management and one for the LFE channel. Bass management is a crossover between a subwoofer and a main speaker so that each covers part of the frequency range. In my case, the bass manager maintains left/right stereo positioning, yielding full-range stereo. It directs low-pass signal for the left front and left surround to the left sub and the low-pass for the

Left front channel frequency response, psychoacoustic measurement, uncorrected (brown), corrected (red). Psychoacoustic measurement uses a continusly-variable FFT window, from very long at low frequencies, to near-anechoic at high frequencies, the way that the ear responds to sound.



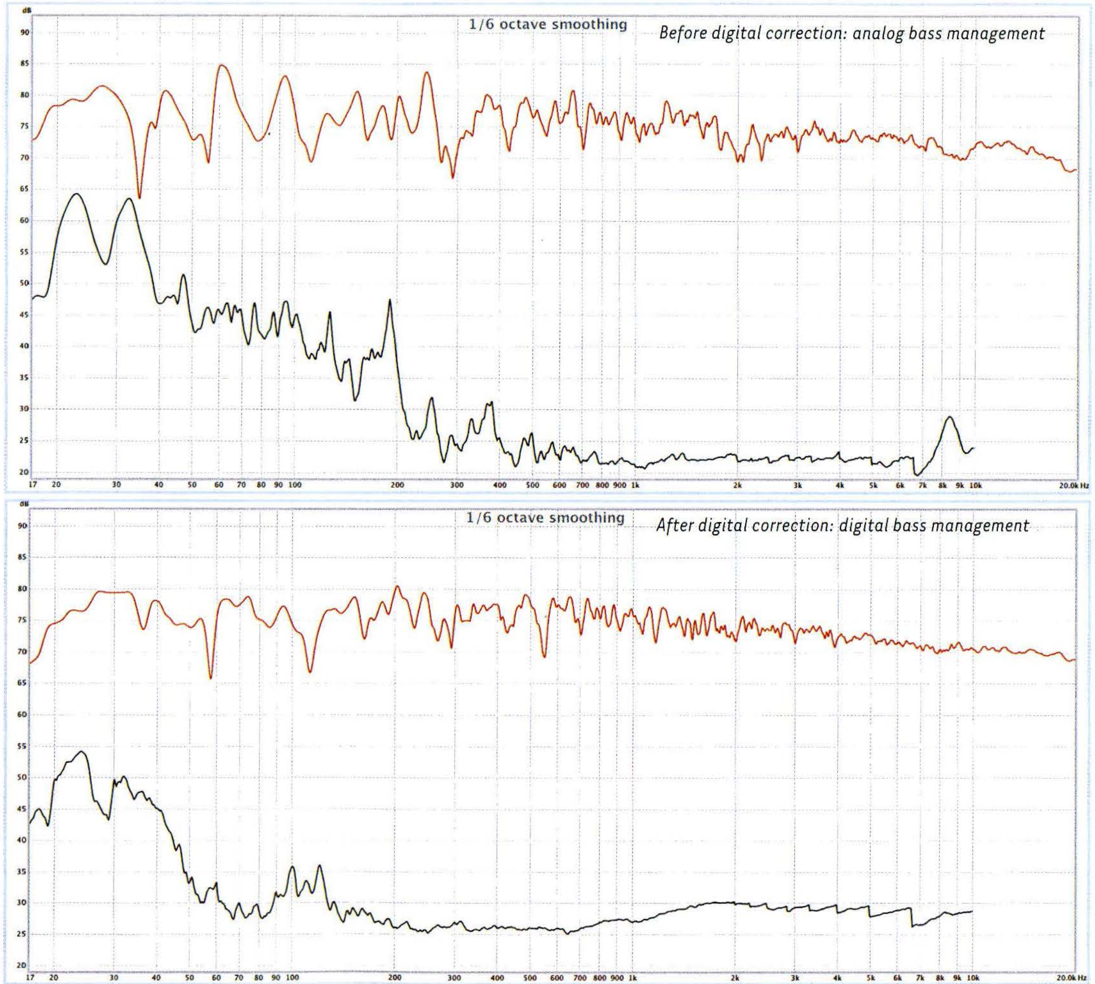
right front and right surround to the right sub. The low-pass for the center is split by the digital matrix and feeds both subs.

We must not confuse the "point 1" or "LFE" channel with the concept of bass management. Strictly speaking, LFE is not part of bass management. LFE is a separate effects channel, originally intended for motion picture sound effects, though many music mixers use it for extra headroom (See Chapter 11). Feeding two subwoofers doubles the LFE output (by 6 dB). This is an advantage, because it takes us closer to our +10 dB goal described below without requiring excessive analog gain. The other 4 dB are made up by analog gain in the subwoofer and the LFE DAC output. We also implement a low-pass filter at 120 Hz in the LFE channel to keep high frequency information out of the LFE reproduction.

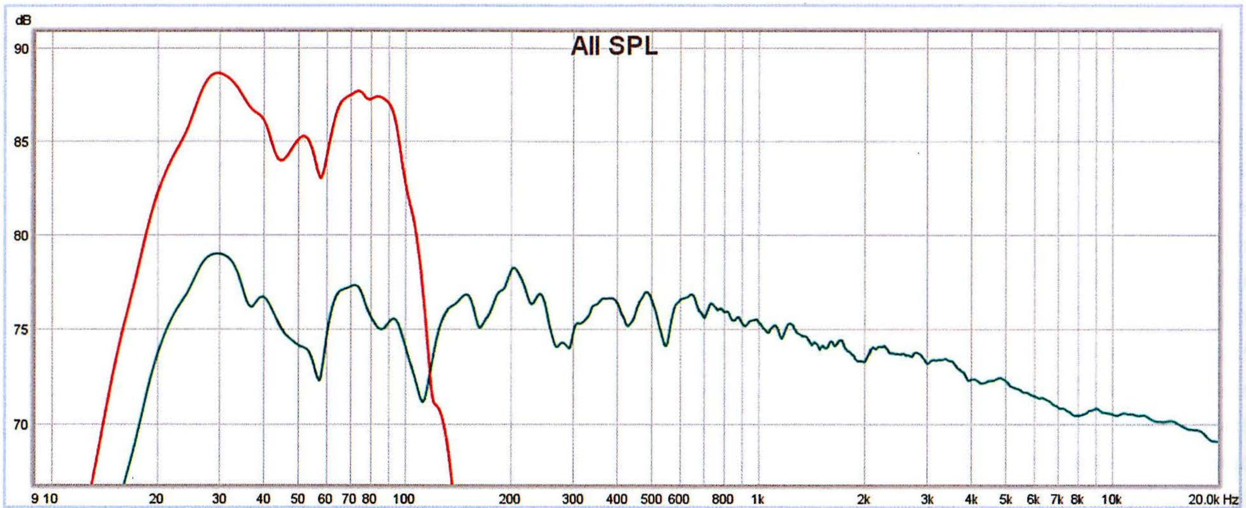
The image on page 287 shows the precision of the phase and impulse response correction. The uncorrected and corrected responses are purposely offset in time for visual clarity. At right (green) is the impulse response of the uncorrected front left channel, which shows the influence, polarity and time spread of each individual loudspeaker driver. The first positive-going impulse is the tweeter, followed by the negative-going midrange and positive-going midwoofer, and eventually the subwoofer, whose amplitude is masked in this plot. At left (red) is the corrected impulse, which is extremely tight with excellent single-direction amplitude, largely positive-going in polarity, and with comparatively little under- or overshoot. The multiway system performs more like a single, crossoverless transducer, with greatly-improved transient and phase response. Sonically it gains the purity of tone, transparency and impact characteristic of a planar electrostatic loudspeaker with the headroom and power of a multiway dynamic! This remarkable improvement is only available with FIR-based digital correction. This display also confirms that acoustic polarity is absolute for all channels with this positive-going impulse.

The correction system uses a psychoacoustic approach[6] to measurement and correction, with an FFT time window whose width varies from very wide at low frequencies to near-anechoic at high frequencies, approximating the way we hear. In other words, at low frequencies, the ear incorporates the sound of the room, but at high frequencies, the ear focuses primarily on the direct sound of the speaker. Next, Acourate performs a transient simulation, to decide which peaks and dips are important to the ear and estimate their amplitude. To date I have not had to manually correct any measurements, unlike other measurement and correction methods I have tried. Pictured (page 288) is the left-front frequency response, psychoacoustically measured, before and after correction. Notice the smooth, extended and linear high frequency response after correction (red curve on page 288). The total system variation relative to the target is about + or -1 dB, an exemplary response. Conservative mastering engineers would be concerned about such radical corrections, but the proof is in the listening. From my point of view, the ragged response of any multiway loudspeaker is an obvious problem that we could never effectively deal with before the power of FIR-based correction.

The system also corrects phase response, reducing phase shifts that would normally be caused by corrective equalization, and the phase shifts present in every analog-domain crossover (greater than first order). Phase shift can only be corrected in the digital domain via FIR filters. Also remember that every analog loudspeaker designer attempts similar fixes, tweaking capacitors, inductors, and resistors until the response is as flat as he can make it. So every traditional



Left front channel frequency response, 1/6 octave smoothing, 500 ms fixed window (brown). Harmonic distortion (black).



LFE frequency response (red). Main channels (green). 1/6 octave smoothng.

loudspeaker has already been heavily-optimized at the factory, yet it still exhibits phase shifts and amplitude errors in the analog crossover. You can see one of those common errors in the brown curve near the 2 kHz crossover point, a very critical frequency region, which has been completely corrected (red curve on page 288). The psychoacoustically-correct narrow FFT window in this frequency region reveals audible frequency response anomalies that would normally be hidden or smoothed by traditional FFT measurements using a single wide window.

I have implemented a digital crossover to the subwoofers that is both steep and linear-phase, yielding flat response in the crossover region and keeping the drivers in their linear regions. The brown curve on page 288 is measured after this digital crossover but before frequency and phase correction. To my ears, these corrections have proved sonically invisible and completely beneficial, unlike analog or digital-domain correction I have tried in the past with other products and techniques.

All small rooms exhibit low frequency anomalies, as shown in the brown curve. This room is already extensively trapped. As I mentioned earlier, it is possible to mess up the important Schroeder curve and the liveliness of the room if applying trapping indiscriminately. I have already found that adding too many "general bass mode" traps has deteriorated the time and frequency response in other regions. It may be possible to ameliorate the low frequency modes with tuned (narrowband) traps (such as the Bag End E-Trap) or bass mode diffusers. I am completely satisfied with the sound of the hybrid solution provided by trapping plus digital correction.

On page 289 are two plots of frequency response (brown) and harmonic distortion (black) measured with Room EQ Wizard. Since REW uses a single FFT window and "standard" 1/6 octave smoothing, it shows a slightly different frequency response than the psychoacoustically accurate response. A wide FFT window masks many mid and high frequency anomalies which can still be heard by the ear. REW can use a narrow window but with only one window per graph I chose wide to be more accurate at low frequencies. The top graph is the old analog bass manager. On the bottom is the new digital bass manager and digital response correction. The new plot shows a 10-15 dB reduction in harmonic distortion, corresponding to a decrease from about 3% distortion to about 0.3%. I believe the startling improvement in harmonic distortion is due to the steep crossover keeping the drivers more within their linear region, taking low frequencies away from the small drivers and high frequencies from the large drivers. This is such a surprising improvement that I should confirm it with another measurement, but unfortunately, the old analog bass manager has been completely removed. Even if the harmonic distortion difference is not as drastic, clearly there has been a significant improvement.

Level Calibration

During the correction process, all five main channels have been adjusted to produce equal level, and the subwoofers have been completely integrated with the mains to produce a total flat frequency response. We have also calibrated levels with the analog output gains of each DAC or loudspeaker amplifier until the narrow-band pink noise signal for each main channel reads 83 dB on the SPL meter located at the listening position with the digital monitor gain control set to 0 dB, which is the calibration point. The meter is set to C-weighted, slow speed. This is the case for all five main channels, but not for the LFE channel.

The LFE is calibrated within its limited bandwidth to produce 10 dB higher SPL than the main channels. This is done in the analog domain to get low frequency headroom without overloading signal in the digital domain. Pictured (page 290) is a 1/6 octave-smoothed frequency response plot comparing the LFE to the main channel.

Phantom Center Check

It's useful to cross-check the system performance with test signals and our ears. We can check the **phantom center** produced by sending an identical signal to left and right front speakers, while listening at the central position. This confirms the front main speakers are in polarity and there are no acoustic anomalies. Play a mono pink noise source panned to the middle (or 2 channels of identical signal) and verify the phantom center appears as a narrow virtual image at the physical location of the center loudspeaker. Since the correction system matches the frequency response and levels of all channels to a very precise standard, the mono center image should be tight and precisely located. We then compare this image to sending pink noise to the center speaker only, and confirm that the virtual image and a real image are located in the same place.

After listening to music, we might want to tweak the angles (toe-in) of the left and right main speakers until the phantom image is better, or if the soundstage does not sound accurate listening from the central position. In a perfect world, if we do change the toe-in, all measurements and corrections should be performed again, since the time delays may have to be minutely adjusted.

Spatial Averaging?

Some authorities recommend spatial averaging to get the best average bass response for all listeners in different seats. With Acourate's transient simulations, I have not found this to be necessary. But there is still one "sweet spot." Plus, the mastering system is meant to be auditioned by one engineer located in one spot.

Subjective Assessment

It's time to listen to some music and confirm that our subwoofers are perfectly integrated with the rest of the system. A properly-adjusted subwoofer should not make itself obvious by its presence, but only by its absence. Listen to music with the subwoofers on and off. They should not sound "lumpy" — they should simply add a sense of weight to the extreme low end. If the crossover frequency is 60 Hz or below, then we may hardly notice a difference except for the additional solidity to the sound. That's the way it should be!

Finding the right recording to evaluate bass is difficult, because recordings of bass are all over the map. An excellent way to evaluate a full-range system is with a naturally-recorded string bass. My favorite test record is one of my own stereo recordings: Rebecca Pidgeon, "Spanish Harlem" on Chesky JD115. This song, in the key of G, uses the classic I, IV, V progression. Here are the frequencies of the fundamental notes of this bass melody:

49 62 73

65 82 98

73 93 110

If the system has proper bass response, the bass should sound natural; notes should not stick out too far or be recessed. Start with the subs and high pass filter turned off and verify the lowest note(s) are a little weak. Then insert the subs and crossover and verify that they restore the lowest notes without adding any anomalies. The addition of the subs should not move the bass instrument forward or backward in the soundstage or become vague in its placement (an indication the subwoofers are too far apart). Once we are satisfied, we take a break and enjoy Rebecca's performance for its natural acoustic reproduction of voice, string and percussion instruments, and the acoustic depth of a good recording hall.

We are off to a good start! Then we listen to various stereo and surround recordings that we are familiar with, to see how the system is performing. Now let's sit back and enjoy our calibrated multichannel reproduction system!

1
Holman, Tomlinson (2010) Sound For Film and Television, Focal Press.

2
Plumbline. From the Latin *plumbum* for lead. Attach a small weight to a piece of string to mark the position of a suspended object.

3
Braasch, Jonas; Martens, William L.; Woszczyk, Wieslaw (October 2004). Modeling Auditory Localization of Subwoofer Signals in Multi-Channel Loudspeaker Arrays. *Journal of the AES* Preprint Number: 6228. Convention: 117 (October 2004). Griesinger, David. Speaker Placement, externalization and envelopment in home listening rooms.

4
International Telecommunication Union, specification ITU-R BS.775-1.

5
Readers can construct their own chart, or obtain a kit with laser chalk and alignment chart from Arcam or one of its dealers. Thanks to Geoff Meads for coming up with the simple but elegant idea of using the "modern technology" of a laser chalk and chart.

6
These papers discuss psychoacoustic analysis and correction of loudspeaker and room systems:
Johnston, James D.; Smirnov, Serge (October 2007). A Low Complexity Perceptually Tuned Room Correction System. AES Convention Paper: 7263. Convention: 123 (October 2007).
Johnston, James D.; Jot, Jean-Marc; Fejzo, Zoran; Hastings, Steve R. (November 2010). Beyond Coding, Reproduction of Direct and Diffuse Sounds in Multiple Environments. AES Convention Paper: 8314 Convention: 129 (November 2007).
Fejzo, Zoran; Johnston, James (May 2011). DTS Multi-Channel Audio Playback System: Characterization and Correction. AES Convention Paper: 8379 Convention: 130 (May 2011).

7
Thanks to Bruno Putzeys for helping refine acceptable distortion specs for a mastering-quality monitor system.

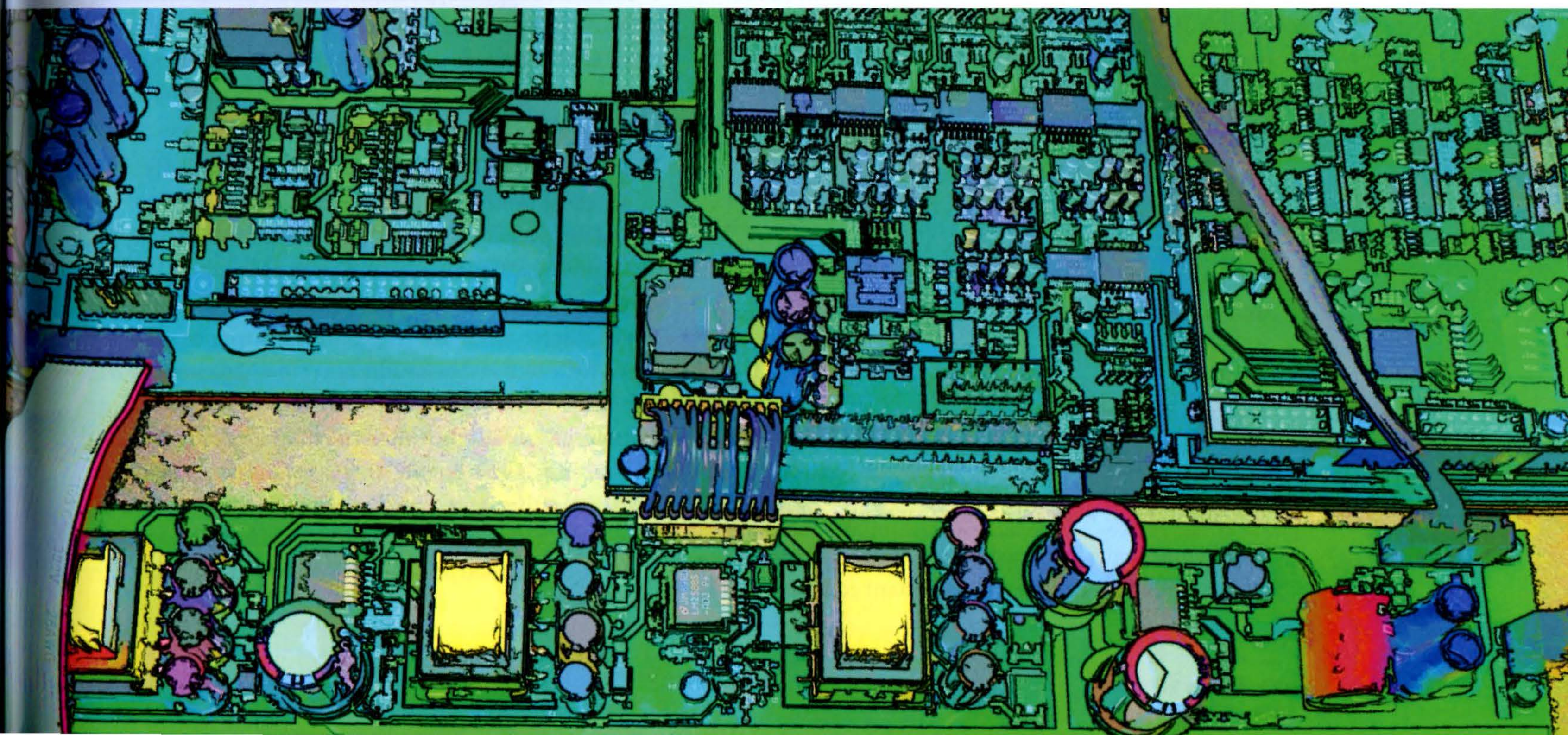
CHAPTER 22

Analog and Digital Processing

I. Introduction

This Chapter tries to reconcile what our ears hear with what our instruments measure. For example, we may measure objective degradation but perceive subjective improvement, or measure objective improvement but perceive subjective degradation! As in the rest of this book, I try to separate fact from opinion, but my biases may show now and then in this Chapter and the following.

Since the first edition of this book appeared, the number of transistors on a typical integrated circuit has roughly doubled every two years — as predicted by Moore's law.[1] That's about eight doublings!



There have been great advances in digital audio processing quality as a result. I once claimed that there is no such thing as a completely transparent audio processor, but I now feel that there *are* a few sonically-transparent audio processors. Even so, I never take a processor for granted: I always listen. Three "transparent" processors chained in series may not sound transparent! Though this seems like a contradiction, it can be explained once we realize that the artifacts of any processor may be "below our perceptual radar," but once several of them have been chained, cumulative artifacts rear their ugly heads.

When processing audio, it's important to consider the tradeoffs. The mastering engineer must be able to recognize when the interests of the client are best served simply by leaving their recording alone — either because the recording is so good it does not need further work, or if the gains due to processing would not warrant the losses due to the same processing.

II. The Ironies of Perception and Measurement

The Greatest ADC I Really Never Heard!

One day I installed a pair of converters in my processing chain for evaluation. I was amazed by their sound. These converters improved the depth and dimension of my sources! It was such an amazing difference that I made a pair of files and sent them to a fellow mastering engineer for his reactions: "Bob, I hear the difference, but guess what, the files that went through the converters are 0.2 dB louder than the originals!" Oops — an 0.2 dB calibration error had slipped into my work. It was a humbling learning experience.

Astoundingly, just 0.2 dB made the sound seem bigger, wider, and deeper on an instant comparison. We don't perceive 0.2 dB as louder per se, but we do hear it as a quality difference. We must conclude from this lesson that loudness is the most important criterion to get right, and not get carried away when we hear tiny differences. The threshold for audible differences could be as low as 0.1 dB. Since loudspeakers themselves vary by small amounts throughout the day due to changes in temperature, expect to hear differences, and learn to adjust to them. Stay humble, and stay alert!

It seems that even in the most professional of studios, the stability and repeatability of your D-A-D chain's level may only be "within 0.3 dB." I can align the system today to less than 0.1 dB and come back in a few months and find it has drifted 0.2 dB. Tiny differences in loading become important in order to retain 0.2 dB consistency. When adjusting DAC gain, I suggest using a Y-cord and a high impedance VTVM so the ADC load is in the circuit when adjusting the DAC output. My precision equalizer raises the level 0.2 dB even when set flat. So in practice, we do our best, we listen, but try not to leap to conclusions.

The Fallacy of Single Number Measurements

We must remember that each measurement only provides a part of the overall picture. Human beings like simple explanations. *Bandwidth* is an example of a single number measurement that is quite misleading. The TC Electronic System 6000 lets the user choose between different low-pass filters for the ADC and DAC. Two of these filters measure "dull" but sound bright! For example, TC's filters that roll off at 16 kHz, called *Natural* and *Linear*, sound more *open* and *clear* than the 20 kHz filter called *Vintage*. The single number, *bandwidth*, tells only part of the story. The missing part of the story is called *passband ripple*, which probably reveals why the *Vintage* filter sounds more closed than the *Linear* (See Chapter 23).

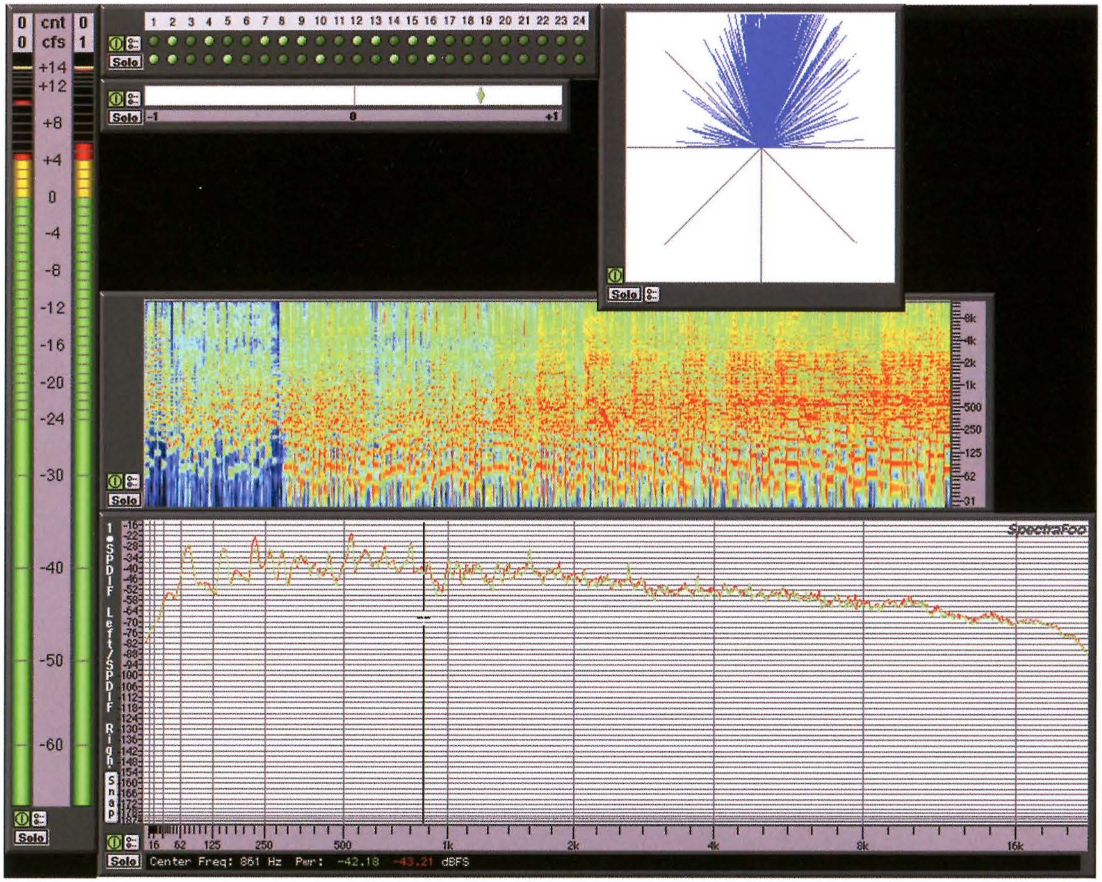
Distortion. Here again humans like to simplify. The most common single number (Total Harmonic Distortion, abbreviated **THD**) measurement means next to nothing. THD is only suitable to confirm that a device isn't broken: if it's supposed to have 0.1% THD and it measures 2%, then something is wrong with it. But otherwise, the sonic difference between 0.1% and 3% is usually not very obvious. The frequency distribution of the harmonics is much more important. We should determine if the distortion is largely 2nd, 3rd, or higher harmonics. 2nd harmonic is easily masked by the fundamental so it's hard to hear. Are the harmonics primarily odd, even, or both? Then measure the intermodulation distortion, especially with digital processors, at 19 and 20 kHz. Keep in mind that the distortion of digital processing is far more bothersome to the ear than that of analog processing, so here the measurements should be orders of magnitude lower to stay below our radar. *Digital processing produces dissonance from harmonic components, which beat against the sample rate, producing inharmonic beat or intermodulation products (aliasing distortion).*[3] The term *inharmonic* means that the type of distortion is not part of the integer harmonic series.

Jitter. Again we tend to simplify. Jitter is usually given as a single value in picoseconds, but without specifying its spectral characteristics. I've observed that the frequency distribution of the jitter is as important, if not more important, than its amplitude. Random jitter sounds more noise-like and less bothersome than signal-related jitter.

Noise. Some analog equipment in our studio has wideband noise floors from as low as -120 dBFS to as high as -50 dBFS (after A/D conversion). However, much of this equipment is perceptually quiet: if I have to put my ear up to the loudspeaker to hear the hiss, I consider it insignificant. A-weighted noise measurements (which are cited all the time) do not seem to correlate well with perception. One particular converter whose A-weighted noise floor is -108 dBFS sounds significantly quieter to me than another converter whose measured A-weighted noise floor is -115 dBFS! This is because the converter that measures better (A-weighted) produces significantly more energy in the region of 3 kHz, but the A-weighted single number does not take the excess at 3 kHz into account. I know very little about the psychoacoustics of noise, but psychoacoustician Jim Johnston says that single-number noise measurements are practically useless; we need to study an FFT display of the entire spectrum, and compare the spikes with the thresholds of the critical bands of the human ear.[2]

"Never turn your back on digital."
— Bob Ludwig

Low bit-rate coding systems. Traditional measurements such as distortion and noise are almost useless for judging non-linear algorithms, particularly coded systems that depend on masking, such as AAC and mp3. Once the ear has been trained to hear their errors, we can easily identify unique digital artifacts that analog technology never produced. One way to expose those artifacts and train the ears is to *transcode*, i.e., to copy signal from one coded medium to another. This is because the noise that was previously masked (imperceptible) accumulates so that on the second coded generation, it rears its ugly head and brings out the



SpectraFoo in action

space monkeys (swishes, gurgles, phasiness and other program-modulated noise). Unfortunately, people who should know better, such as broadcasters and cell phone carriers, transcode all the time. Telephone and broadcast audio sounded much better before the invention of the codec! Broadcasters frequently start with AAC sources, then recode them to broadcast with another codec. Suddenly all the noise and distortion becomes evident. Instead, they should broadcast CD originals; broadcast sound quality will immediately get better.

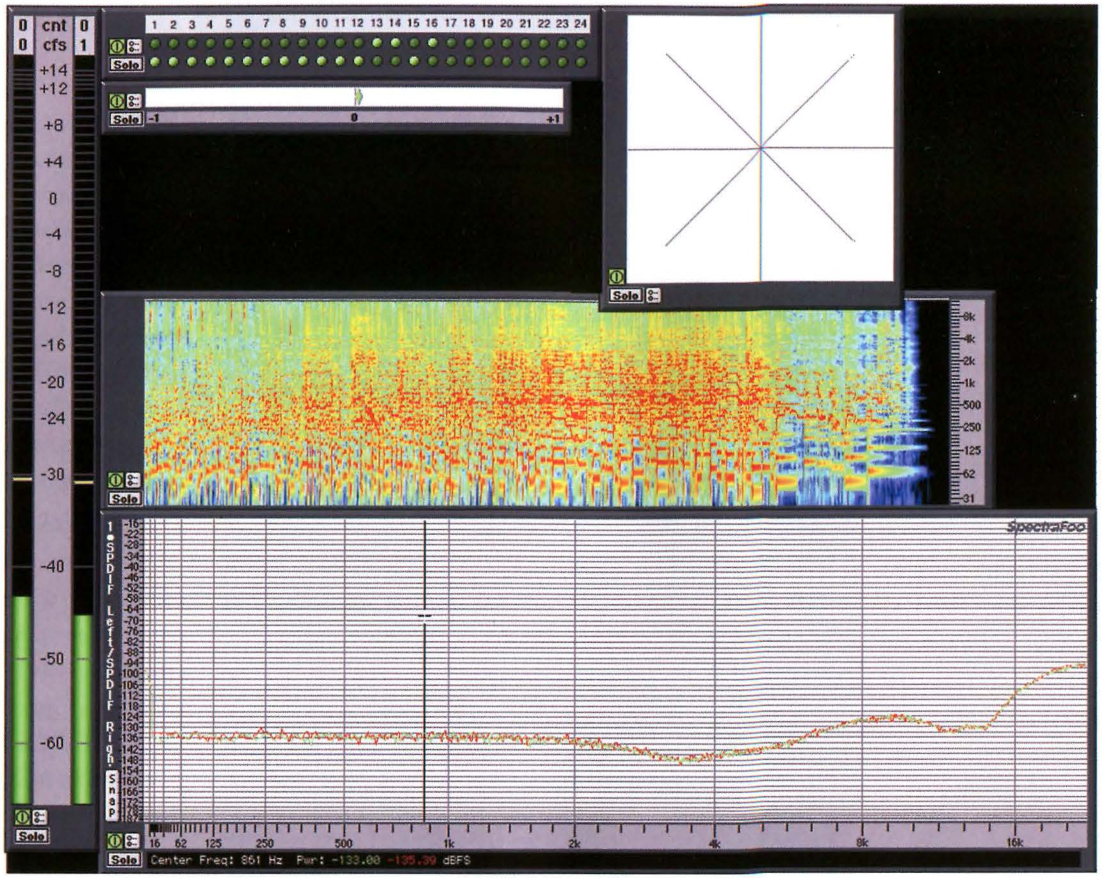
III. Measurement Tools We Can Use While Mastering

FFT: Eye Candy or Real Help?

FFT stands for *Fast Fourier Transform — a mathematical tool which enables us to move between the time (waveform) and frequency (spectrum) domain*. High-resolution FFT analyzers are now very reasonably priced, and they can provide an *early warning system* to protect us from the manifold and varied bugs of digital audio. They're no substitute for the ear, but are a great supplement. For example, pictured here is a screen shot of SpectraFoo™ in action during a mastering session.

At the middle top is a bitscope, currently showing 16 active bits, an indication that the dither generator is probably doing its job. Since one of the symptoms of a dysfunctional processor is to toggle unwanted bits, or hold some bits steady when there is no signal, bitscopes can reveal if the DAW or some digital device is malfunctioning. They can also show if there are any truncations caused by defective or misused processors, though they cannot tell us the degree of distortion introduced by a processor, or whether idle bit noise is significant or simply random. At top right is a stereo position indicator, at a moment when the information is slightly right-heavy. If you use this indicator at all, let it be a visual confirmation of what your ears are already telling you, because the meter has no idea what the music is trying to do at that point in time. Just below the bitscope is a correlation indicator, which reveals that the material is significantly monophonic (See Chapter 9). We use our ears to confirm that the image is not "vague," and perform a mono (folddown) test to make sure the sound is mono-compatible.

At mid-screen is the spectragram, showing spectral intensity over time. This can be useful to identify the frequencies of problem notes, or simply to entertain visitors! At bottom is the spectragraph, whose general rolloff shape gives a rough idea of the program's overall timbre. And I mean rough, because timbre is genre- and program-specific. I advise using ears and a trusted set of monitors.


SpectraFoo during a pause in the music

The figure on this page shows SpectraFoo during a pause in the music. This is perhaps the most useful diagnostic: The bitscope shows only the bottom four bits are toggling, indicating that a high order noise shaper is probably in use (the higher the order, the more bits are active). If this were a movie, you would see four random dancing green lights in bits 12-16. The spectragraph shows the curve of the dither noise, which we can identify as POW-R type 3 or a similar high order curve. Using this analyzer, we can often determine the type of dither used by the mastering engineer on recorded CDs. The correlation meter fluctuates very slightly near the meter's center, showing that the dither is uncorrelated between channels (random phase), which is a good thing, because uncorrelated dither helps to preserve the stereo width.

Wavelab, RME Digicheck, and Sequoia provide similar visualization tools, but SpectraFoo is the king of resolution and color visualization. During mastering, my only tool is a loudness meter and sometimes a correlation meter that I engage if I want to confirm what my ears are already telling me. When capturing to CD, I add a bitscope as visual confirmation the dither is in operation. I may add some of these other visual tools on occasion.

PLR Reveals Equipment Performance

In Chapters 16 and 18 I discussed PLR (Peak-to-Loudness Ratio), which is a useful measure of program transients and an indicator of differences in sound quality. One day I inserted my Pendulum ES-8 in the analog chain at unity gain, just to provide character (not even compressing). It warmed up and added the right color and glue for a rock recording that had been mixed entirely digitally (we say "in the box"). I noticed that the PLR of the signal going into the ES-8 was 2 dB greater than what was coming out, so even though it was not compressing per se, the ES-8 reduced the PLR of the original recording by 2 dB. This is due to a combination of the ES-8's input transformer and tube circuitry, which compresses the signal gently and invisibly even without threshold or gain reduction — thus reinforcing Dick Pierce's adage: "Distortion is compression is distortion is compression is distortion is compression."

IV. Measurement Tools to Analyze your Equipment

It helps to be fluent in the use of diagnostic tools to measure equipment performance. I frequently use SpectraFoo to measure and confirm the performance of a piece of gear, as well as the Audio Toolbox. When I can afford it, I'd like to own a dedicated analyzer like the Audio Precision or the Stanford Research. Keep in mind the fallacy of single-number measurements: measurement tools are much better at finding faults in gear than in gauging how good it may sound. As a preventative measure, we can analyze our equipment by sending test tones into the processors and observing an FFT. We can measure the distortion and noise of our tube processors when they are new, compare and do a checkup every six months to see if the tubes have deteriorated.

An FFT can confirm if the bypass switch on a digital processor is truly working. SpectraFoo can measure distortion 40 dB below the 24-bit noise floor! This is because it is splitting the energy into smaller segments in the frequency domain. So we can compare the distortion and noise of processors that simply truncate at the 24th bit with others that use a longer internal wordlength and then dither down to 24 bits. When connected digitally, interface jitter (see Chapter 23) is completely irrelevant to FFT analyzers, which strictly look at data.

The Gonger — A Great Listening Test

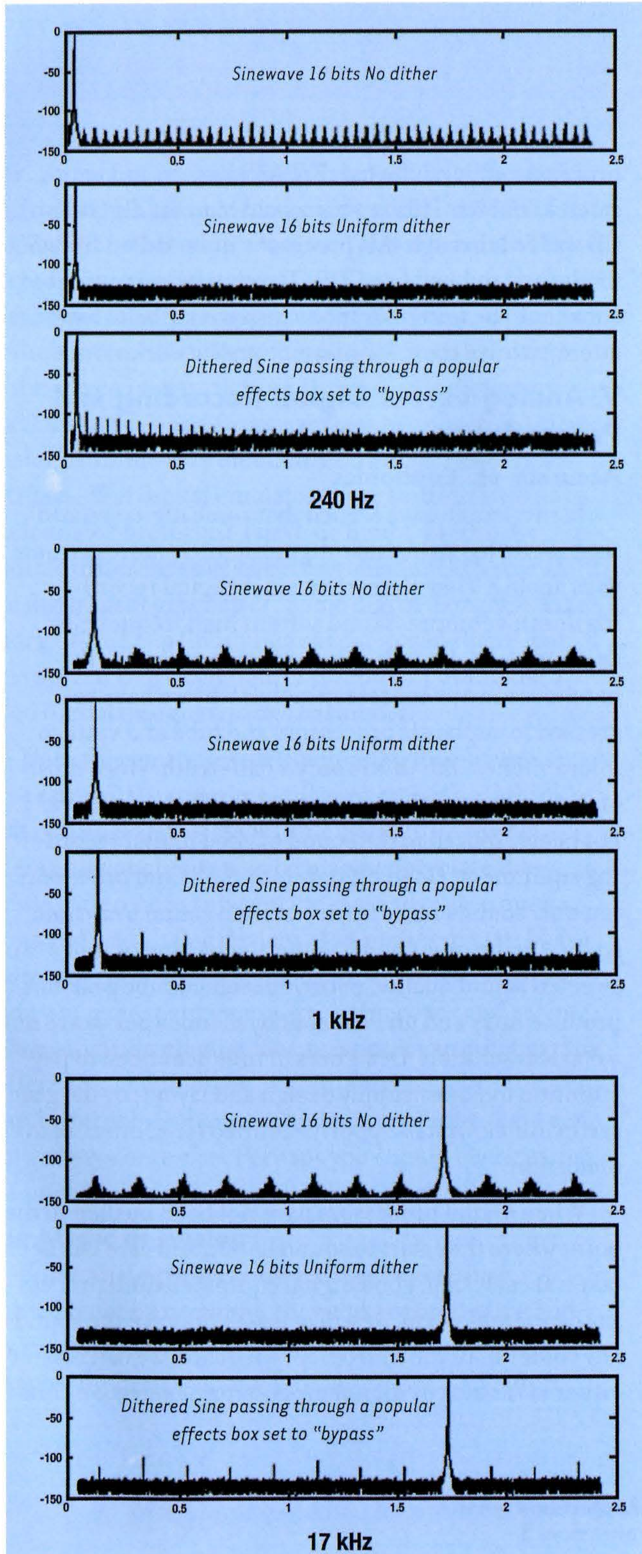
Steady-state (static) sinewave measurements are misleading when measuring nonlinear processors like compressors. A more effective judgment can be made simply by listening using the *gonger* (aka *bonger*), originally developed by the BBC's Chris Travis and available on a test CD from Checkpoint Audio (see links). This powerful listening test is a sinewave that modulates through various amplitudes, in the process exercising and revealing any amplitude non-linearities in the signal path. We play the gonger through the device under test and listen for noise modulation, buzz or distortion. I run the gonger through my system every time I repatch, because it instantly reveals subtle distortion that may be due to clocking errors or bad connections — distortion that may be masked during normal musical passages. The gonger also mercilessly reveals speaker rattle and sympathetic vibrations in the room.

Identity Testing — Bit Transparency

A neutral signal path is a good indication of data integrity in a DAW. Any workstation that cannot make a perfect clone should be rejected. The simplest test is the identity test, or bit-transparency test. Set the device under test to flat and unity gain, then see if it passes signal identical to its input. This is done through a **null test**: We capture the output of the device to a second file and play the two files synchronized, inverting the polarity of one and mixing the two together. If there is any output, the two files are not identical. Some people scoff at an identity test, since analog equipment could never produce identical output. But this test is important to identify digital distortion-makers. The bitscope (or a plugin such as *Bitter*), is another verification of bit-transparency: a device is likely bit-transparent if we selectively put in 16 bits, then 20, then 24, and get an identical result. We can also watch a 16- or 20-bit source expand to more bits when the gain changes, during crossfades, or if any equalizer is changed from the 0 dB position.

The Sound Effects of Defective Digital Processors

Never take a digital processor, or any DAW or computer that processes audio, for granted. The mouse is a dangerous weapon! Or, when software is changed or updated, we should not assume that the manufacturers have found all the bugs and we should assume that they may have created new ones. We even need to ensure that BYPASS mode, which seems seductively simple, actually does produce true clones in bypass. The illustration (pictured at right), courtesy of Jim Johnston, is a series of FFT plots of a sinewave, showing the type of non-linear distortion products generated by truncation without dithering. The top row is an undithered 16-bit sinewave. Note the distortion products (vertical spikes at regular intervals, not harmonically related to the source wave). The second row is the same sinewave with uniform dither. Note that the distortion products have disappeared. The bottom row is the formerly dithered sine-wave, going through a popular model of digital



MYTH:
It's a digital processor, so there's no generation loss.

processor with a defective BYPASS switch, and truncated to 16 bits. This is what would happen if a (16-bit) CD was fed through this processor in so-called BYPASS mode, and dubbed to a CDR! That is why every processor should be tested for bit-transparency before we attempt to use them for master-quality work.

V. Analog versus Digital Recording and Processing

Accuracy vs. Euphonics

Many people have argued that some digital recordings sound harsh because digital audio is more *accurate* than analog. They claim that, since digital recording doesn't compress (and soften) high frequencies as analog tape does, this *digital accuracy* reveals the harshness in our sources, which is why we have regressed to euphonic processors and tube and vintage microphones. But this is only a half-truth, since most of these arguments come from individuals who have not been exposed to the sound of good digital recording equipment. Good digital equipment and processing not only sounds accurate, it can even sound *warm and pretty*. On the other hand, while analog tape or tubes can sweeten sound quality, poorly-designed tube gear can produce fuzzy and unclear sound, so tubes per se are not a magical solution. Designers of tube gear need to pay attention to power supply design and layout, or the gear can exhibit crosstalk, poorly-defined bass, or excessive distortion.

When digital processors have not been pushed to the point where they start to sound harsh, and care has been taken at each step, good digital equipment and processing can sound warm, or will at least not make the sound any colder than the source. So, with care, we can master either via analog or digital processing or selective combinations of the best of both worlds. I tend to use digital processing when transparency is needed, especially if a mix has already passed through analog tape or previous analog processing. You can easily have "too much analog" or "too much digital" in any given project.

Avoid cheap digital equipment, which is subject to edgy sounding distortion with any number of causes:

- Sharp filters
- Low sample rates
- Poor conversion technology
- Low resolution (short wordlength)
- Poor analog stages
- Jitter
- Improperly-applied or misapplied dither
- Clock leakage in analog stages due to bad circuit board design
- Induction or ground loops caused by placing sensitive A/D and D/A converters inside a computer chassis with all of its interference or simply poor circuit board layout. It takes a superior power supply, circuit board layout and shielding design to keep a converter inside a computer sounding good. Most external converters sound better than those placed inside of computers.

Using a high quality D/A/D chain without any processing at all can be one of the most subtle "analog processes" to use. Years ago I would never have dreamed of such a thing, because the degradation of a D/A/D loop was enough of a loss to outweigh sending a signal through it. But today, a good 96 kHz D/A/D loop can subtly "soften" sound, and the best converters are barely perceptible. Recently I found a converter pair that seems audibly transparent at 96 kHz. It's amazing how far we have progressed.

When considering digital processing, numeric precision and resolution (internal wordlength) are important. One difference between analog and digital processing is that noise in analog gradually and gently obscures ambience and low-level material. It is random (uncorrelated with the music) and does not add distortion at low levels. By contrast, numeric imprecision in digital processors causes noise-like errors that increase at low levels, and are correlated with the music. When it affects the body and purity of a mix, because of the addition of inharmonic distortion, it produces an edgy, colder sound, which I call **digititis**. So, depending on the quality of the digital processing — and the number of passes through that circuitry — it might be better to mix through a high-quality analog console or processor. But I've noticed considerable improvement in all-digital mixes received for mastering, not only because the equipment is getting better, but also because engineers are learning to avoid the pitfalls of digital (e.g. having too many low-resolution plugins in the signal path, pushing them too hard, or not dithering the signal).

Digital Progress

The digital situation keeps getting better each year. If we choose digital, we must be aware of the aforementioned weaknesses and distortion mechanisms. Moore's law and care on the part of designers will soon overcome these obstacles.

Digital Emulators of Analog Circuits

When I said that digital consoles and DAWs can either preserve sound quality or make it colder, there is a third option: analog emulation. There are now many digital circuits and plugins which purport to add analog-style warmth or color to the signal chain. A few sound excellent, some sound quite good, others are fair, and many are quite poor. It helps to run these at higher sample rates. Measure if possible, and look for inharmonic distortion. One good test is to send a very high-frequency sine wave at high level through the processor and look for aliases on the FFT. Then, with music as the source, listen for symptoms such as shrinkage of the stereo image, digititis (inharmonic distortion and harshness), and fuzziness (loss of definition). The analog units they are emulating do not produce these artifacts. Will digital emulators ever equal the sound of their analog brethren? I have no doubt, eventually. The audible differences keep getting smaller each year as the digital gear gets better. Some digital designers take shortcuts to keep their plugins cost-competitive, but when given free reign, lots of computer power, and a good digital designer, *watch out analog!*

Does this mean we can finally abandon our analog processors? My answer is still a vehement *no.* I think that real analog processors still matter a lot to engineers and our clients who hear and appreciate the subtle differences between the most advanced emulations and real analog devices. As a matter of fact, I have increased the size of my analog processing arsenal at a time when most engineers are cutting back and relying increasingly on plugins. The reasons: I can still hear the difference, and in many cases the analog gear provides a wonderful color, character and purity of tone that I cannot get anywhere else. Perhaps you should check with me once Moore's law has octupled again!

The Magic of Analog?

Some analog processors are *magical* because, although not transparent, they add an interesting and exciting sonic character to music. As mentioned in Chapter 6, the classic analog compressors' signature

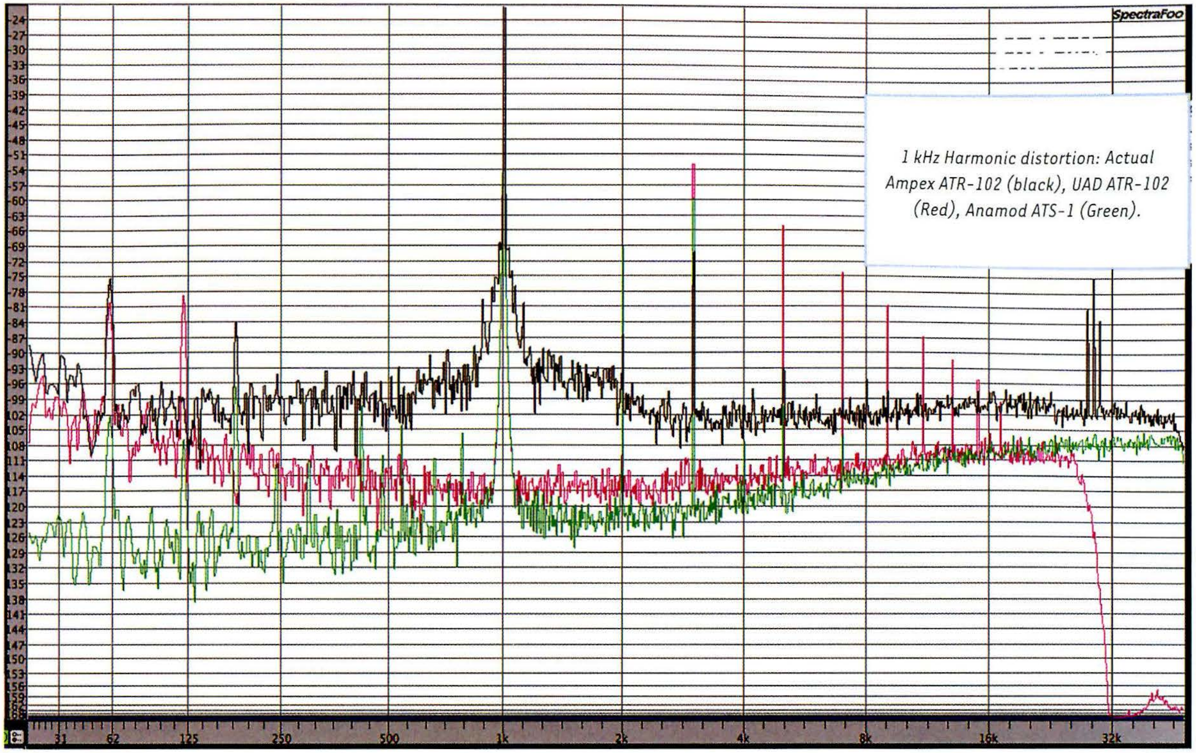
MYTH:
It's a digital console. It must be better than my old analog model!

advantages come from a unique combination of attack and release characteristics (which can be well emulated digitally), zero alias distortion (a type of distortion that occurs only in the digital model), and some degree of integer harmonic distortion (which is difficult to emulate well). Static distortion measurements don't tell us every reason why some compressors sound *excellent* and others hurt our ears. Certainly the Weiss digital compressor does not sound *digital,* so we know it can be accomplished with programming skill and expensive DSP. But it does not achieve as much punch or warmth as my analog compressors, perhaps because of its lack of distortion. As my skill at operating the TC Electronic MD4 improves, I often get the sound I want in the digital domain, supplemented — when the recording needs it — by the warmth of a nice analog tube stage used simply as a pass-through. This gives us the best of both worlds: precise control with digital processing, combined with the genuine warmth of an analog processor. It is possible to cover (mask) some dissonant distortion (if it is low enough) by adding sufficient consonant distortion to the chain. This is probably why audiophiles like to add tube preamps to an otherwise solid-state chain.

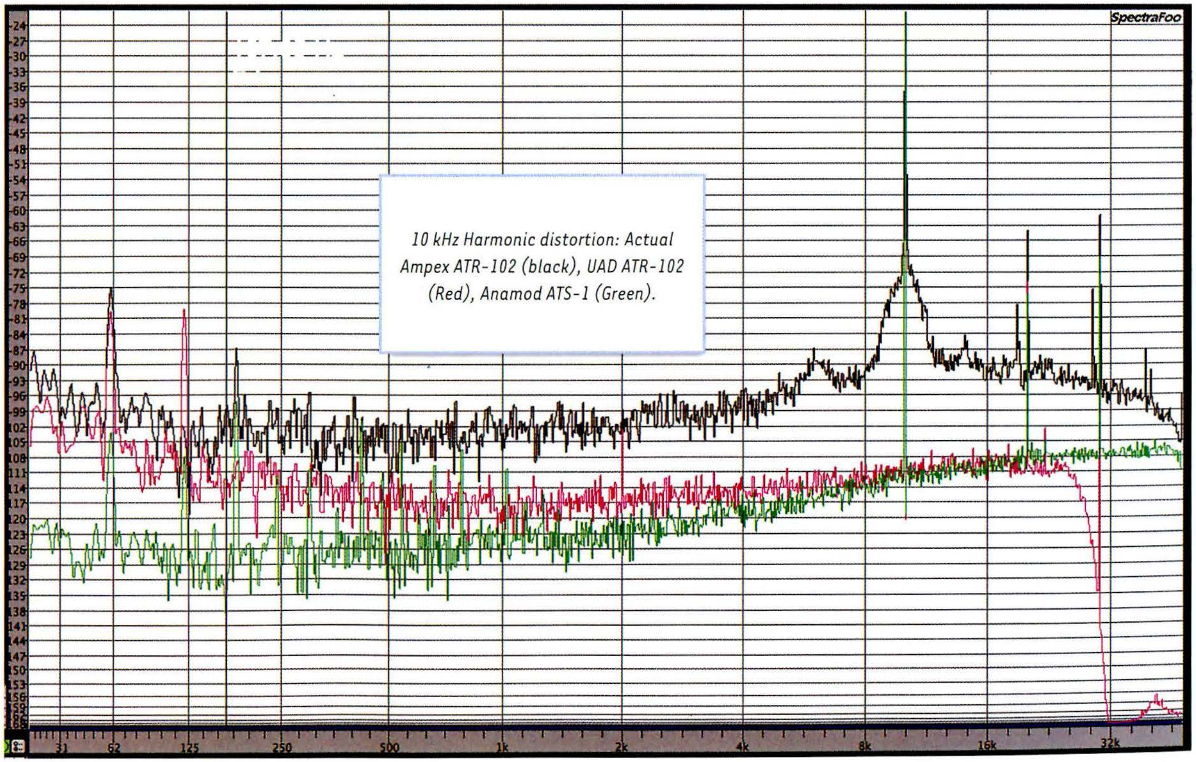
Analog Tape and Emulation. Analog tape recording is a perfect example of a process that objectively measures worse, but subjectively is often desirable. 15 ips with Ampex 456 is considered to be the sound of rock and roll by many engineers. Noise may be another reason why analog tape sounds more *musical* to many people. Maybe -120 dB or even -70 dB noise is better than -144, just enough to cover the ugly parts of the distortion of even some of our best analog and digital gear, so perhaps it's good that our converters aren't any quieter. In addition, noise-free recording media can sound very *sterile* because the nits, cracks and distortions caused by the musicians and their amplifiers are revealed by the quiet media: another case where accuracy is not necessarily desirable, and where extra noise can be euphonic by masking less desirable noises or distortions. We must remember to consider noise-masking as a tool — controllable noise is available in the Anamod and UAD analog tape simulators.

Pictured (page 303) is a measurement taken at 96 kHz sampling, comparing harmonic distortion of "the reel thing" versus two forms of analog tape emulation. One clear difference between the real tape and the emulators is the tape noise. I could have turned on the noise of the emulators, but we probably would not learn any more from having it on. Looking first at the 1 kHz test, it's interesting to see that both Anamod and UAD consider the second harmonic to be more important than the real thing. The red trace (UAD) at 2 kHz is hidden behind the green and about 3 dB lower than the green. Both second harmonic traces of the emulators are 12 to 15 dB higher than the real thing. Perhaps the machines which UAD and Anamod used as models had more second harmonic distortion. Or perhaps RMG tape has less second harmonic distortion than Ampex, or the overbias reduces it. In any event, I do not consider second harmonic to be the key to analog tape's magic: I think it is the third harmonic. We can see that the real tape machine's third harmonic is about 9 dB lower than the Anamod's and 15 dB lower than the UAD's. So both emulators may sound a little bit fuzzier or richer than the real thing! The UAD has considerably more fifth harmonic than the real thing and the Anamod has none (I confirm there is no hidden green line at 5 kHz). Potentially disturbing are the extra harmonics the UAD produces above 5 kHz, which I believe to be unintentional artifacts of the DSP. Being higher in frequency, they are likely subtle. Only your ears can make that judgment. And notice that the real ATR is not without hidden secrets, i.e., the supersonic noises in the 30 kHz range.

At 10 kHz, the real ATR and the two emulators produce second harmonics very much in the same measured ballpark, but since the second harmonic is 20 kHz at levels 30 or 40 dB below the threshold of hearing, it's probably an academic observation, unless intermodulation distortion (not measured) causes other problems. The real ATR produces an intermodulation product at 18 kHz, which is measurable but also well below audibility! It is interesting to see the Anamod is faithful to the real machine all the way to the third harmonic (30 kHz), but the UAD is not capable of doing that, and appears to roll off above 20 k. However, the higher the test frequency, the fewer the extra unwanted harmonics produced by the UAD. In summary, I prefer the sound of the Anamod, but both simulators are excellent devices, and only well-tuned ears will be able to hear their subtleties. Emulators have come a long way: I wouldn't kick these out of bed!



1 kHz Harmonic distortion: Actual Ampex ATR-102 (black), UAD ATR-102 (Red), Anamod ATS-1 (Green).



10 kHz Harmonic distortion: Actual Ampex ATR-102 (black), UAD ATR-102 (Red), Anamod ATS-1 (Green).

The Reel Comparison!

The analog tape machine is a well-maintained Ampex ATR-102 courtesy of engineer John Chester and Steve Puntolillo, owner of Sonicraft. The Ampex uses a Flux Magnetics repro head, 1/2" tape running at 15 IPS, with RMGI SM911 tape, biased 3 dB over at 10 kHz, flux level 285 nW/m. The analog emulator, an Anamod ATS-1, is set for ATR-102 and Ampex 456 at 15 IPS, 250 nW/m. The difference between 250 and 285 nW/m is only 1 dB so this is a pretty fair level comparison. The digital emulator is a UAD ATR-102, set the same as the Anamod. For the two emulators, I turned off tape noise and set bias and EQ to "normal". The UAD is set to 1/2", noise is off, crosstalk on, transformer emulation on, and wow and flutter is set to off. In my experience with Jamie Howarth's *Plangent Process* for removing wow and flutter, W&F is an objectionable-sounding error. I haven't found an analog tape that doesn't sound better with the W&F removed via Plangent, so I'd prefer that it be set to off. Of course the Anamod cannot simulate W&F and the real ATR's W&F is manifested by the wider skirts at the bottom of the fundamental black trace.

Analog Purist?

Does knowing what's under the hood help us to make a judgment about an analog processor? Shouldn't we just listen and form a judgment to avoid our biases? I purchased a digitally-controlled analog equalizer that I know uses MDACs as its automated-control ingredient. Generally I am concerned about the distortion caused by MDACs, as I have had bad sonic experiences with several devices that employ them for remote control. I doubt that harmonic distortion is a valid method of judging MDAC distortion, because they usually measure low enough (below 0.3%) when employed properly, but for me most MDAC-equipped boxes add a veil over the sound, and sometimes a harshness. One device I auditioned sounded initially "nice and warm," but it turned out to be just fuzziness masquerading as warmth. I eventually found a test signal that revealed its issues: plain ol' pink noise, whose peaks seemed to overload the internal circuitry and produced "clicky" artifacts even at low input and output levels. Most music masked the processor's obvious artifacts, which manifested as "fuzziness." I sold that EQ before too long.

Still, I'm very interested in digitally-controlled analog gear. I think it is the future of analog technology in an increasingly fast-paced world, where we have to perform rapid revisions and resets. Our customers demand that we do revisions quickly, which requires resetting our myriads of processors. I purchased another digital EQ that happens to contain MDACs and was pleased to discover it sounds transparent, pure, and "analog," without any of the fuzziness I had heard in previous MDAC-equipped gear. This exception "proved" the rule. A new electronic volume chip in the Muses series from NJR is a very promising replacement for MDAC and is beginning to be used in high-end consumer gear. My conclusion is that, when purchasing automated analog gear, be skeptical but keep an open mind! I don't have time here to enumerate all the technologies that enable analog automation, but I can assure you there are differences in sound quality and transparency due to the technologies and also to the skills of the developer. I feel that our test methods for evaluating technologies have to evolve to expose the distortion-producing mechanisms in the technologies, especially with dynamic or varying signals (as in real life!). One of the best signals to try is Jim Johnston's complex buzz signal (see links), which may reveal issues that single or dual-frequency measurements do not.

The Summing Amp Controversy

The biggest audio snake oil being sold today is the dedicated analog summing amplifier, when it is advertised to avoid so-called *problems* in digital summing. It is false advertising to market an analog summing box as "fixing a problem" (with digital summing) that does not exist. So if the problem does not exist, what is the attraction of these summing boxes, and should we use them?

Many analog engineers who have converted to digital mixing complain that their digital mixes lack separation and depth; then they discover that adding analog summing boxes enhances depth and apparent separation. This is quite true (at least for some of the models of summing boxes). But let's get the facts straight: there is absolutely nothing wrong with digital summing: it is essentially perfect, especially since adding numbers is the easiest thing you can ask a DSP to do — equivalent to adding voltages in the analog domain. Digital summing does not have crosstalk or phase shift, and it does not add distortion, as long as the calculations are properly dithered. So, what's the fuss? As a mastering engineer, I've discovered that some analog processors do appear to enhance separation and depth. Since these are just 2-in-2-out equalizers or compressors, the mechanism must clearly be distortion — some combination of phase shift or saturation in their input transformers and the characteristics of their active circuitry. Analog gear is often desired for its ability to enhance depth; so far I haven't found a digital emulator that quite equals the wonder of passing signal through my API 2500 or my Pendulum ES-8, or any number of other fine pieces of analog gear that I use.

Let's face the music: since a simple 2-in-2-out analog processor enhances depth, it's logical that analog summing boxes containing the same components would do the same. Experiments that I have performed have confirmed that hypothesis. We did a shootout of three mixes:

1) in the box (digital mix)
2) mixed through model A of analog summing amp
3) mixed through Model B

Stereo stems were used, pan pots set full left and right, and gains were carefully matched using test tones. Polarity was confirmed as non-inverting in each circuit. The results: Model A summing amp was so pure in its design and its distortion so low that it sounded almost indistinguishable from the digital mix — nice, but somewhat "boring" (it was not a great digital mix). However, Model B, which uses transformers and highly-praised discrete opamps, opened up the sound. It seemed to enhance the separation, depth, and even the definition. I am quite confident that the reason for this sonic change was *not due to the summing*, because the next experiment I performed was to take the entire digital mix (just two channels), and pass it through just two inputs of Model B, as a "pass-through" unity-gain processor — and guess what! It sounded very much like what had been accomplished previously when using all 16 inputs of Model B as a summer. We can only conclude that summing is not the root of the mix engineer's issues. We wonder if manufacturer A was motivated to make its summing box because someone told them there was a problem with digital summing, but they didn't succeed because they made a transparent summing box. Let's just admit that an *ordinary* in-the-box mix craves the coloration of wonderful analog circuitry. However, an *extraordinary* in-the-box mix may not need any analog help. As we can see, it needs to be a particular kind of analog circuit, one that supplies just the right amount and kind of distortion and coloration. We also discovered that a transparent analog chain plus medium-class converters can sound worse than the all-digital mix! Here's a summary of my conclusions:

"Audio processing is the art of balancing subjective enhancement against objective degradation." — Bob Olhsson

- Analog summing can sound virtually indistinguishable from digital summing when transparent analog components and converters are used. In this case, the analog summing (mix) provides no perceived benefit, so what's the point?
- *Any* audio source can gain depth and apparent separation when passed through certain analog components, due to their "friendly" distortion. This is a pleasant "bonus" (or artifact) of the analog chain. Although measured separation may even be less, psycho-acoustically the distortion appears to increase separation and depth. In other words, an analog device

> *"There is absolutely nothing wrong with digital summing other than that it may sound boring."*

can create character which a technically-perfect digital device or mix may not be able to match.

- Too much distortion is as bad as too little: a mix can quickly sound fuzzy and lifeless with the wrong analog components.
- Only the most superior D/A/D chain has high enough converter transparency to justify the subjective improvements of the analog processors. *A poor set of converters reduces separation and depth.* To reduce the compromise, expect to spend as much for a superior 2-channel converter as you would have put into 8 medium-quality channels. Thus the economics of a multichannel analog summing box dictate that the converters employed are likely compromised.
- Large analog consoles are fun and have wonderful character, but here's the rub: There is no need for the mix engineer to invest in multiple tracks of analog summing and multiple D/A/D stages when a single high quality D/A/D stage coupled with a selected 2-track analog module can achieve the same result! Furthermore, I recommend this analog stage be postponed until the mastering, when it can be integrated and fine-tuned in conjunction with other mastering processing which affects the sound.
- It is conceivable that an analog console may sound a bit more "desirable" than a single 2-channel module because of the complex crosstalk and leakage between channels in the analog console. Regardless, I feel this is a small contribution to the sound; assuredly, just 2 channels of the right analog gear adds a nice amount of depth and desirable coloration.
- Compared to an analog mix, a digital mix may appear to have a smaller soundstage. This is not a technical defect — the digital mix is just being accurate! Note that an improperly dithered digital mix is a technical error which would shrink the soundstage.
- Not every style of music benefits from analog coloration, either during the mix or mastering. A lot of musical styles (such as classical music and much jazz) are looking for a "clean" or "transparent" approach.
- Before leaning on analog character to create a semblance of depth, there's a lot that can and should be done through better mixing techniques. During digital or analog mixing, it is possible to obtain good separation and depth using techniques described in Ch. 10.
- During mastering, you can enhance depth in the digital domain using specialized digital processors.

Separation and Timing In Digital Audio

Digital audio presents no obstacle to channel separation or timing resolution. Digital channel separation is literally *infinite*. On the analog side it is limited by noise, so measured channel separation might be 100 dB with good analog gear. Even 20 dB is very good separation; the LP phonograph was lucky to have 15 dB yet sounded quite good in stereo.

It is a myth that CD's timing resolution is inadequate. In fact, 16 bit/44.1 kHz audio has timing resolution much finer than the human ear, as illustrated in the figure on page 307 (courtesy of Dick Pierce). It can capture the difference in timing of two waves offset by less than a sample period. Timing resolution of 16-bit/44.1 kHz is the sample period/number of levels: 22.7 µs/65536 = 346 picoseconds — more than a million times smaller than the capability of the human ear to detect timing differences.

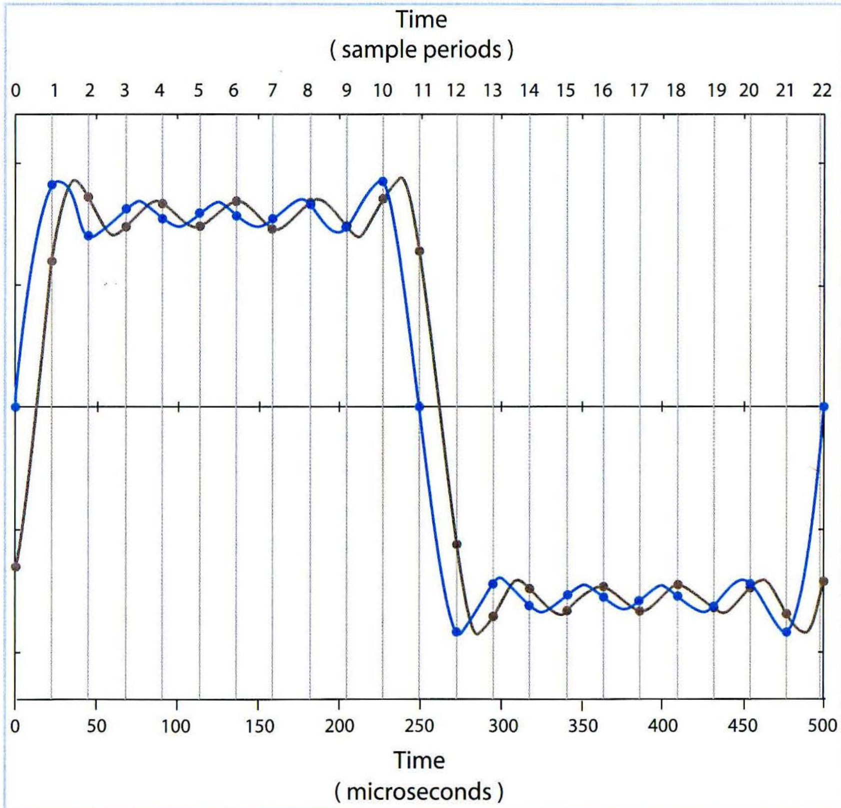
Can You Hear Truncation?

In the analog era, it was pretty easy to push buttons and not get into too much trouble. At least when we did get into trouble by overloading or distorting analog sections of our consoles or preamps, the distortions were fairly consonant. That's the nature of analog. But digital audio errors are *inharmonic* and dissonant, and insidious, for we must be technically knowledgeable in order to avoid these errors: analog audio was easy, but digital audio is hard! No matter how easy developers try to make their software, we cannot get away with digital audio ignorance. For example, someone once asked me if their DAW automatically dithered to 24 bits. My answer was, "I don't know, why don't you test it!" I doubt that any DAW currently on the market dithers automatically without user intervention: there are too many unpredictable variables.

Digital overload distortion is pretty obvious even to the untrained ear, but another form of distortion, truncation distortion (see Chapter 15), is quite audible once you have learned to identify its artifacts. Most engineers do not deny the audibility of truncation at the 16th bit, though some have tried to argue that because the phenomenon occurs at -96 dBFS or so, it must be inaudible. I guess they don't listen for a living. As I've already pointed out, while the dither noise at -96 dBFS is inaudible (at normal monitor gains), many of the products of truncation distortion are well above -96 dBFS, and audible. A simulation with a single tone does not tell the whole story, because especially in a mixer with many tracks, distortion products interact with each other to produce more severe inharmonic distortion products,[4] and distortion accumulates. In most cases, truncation is not heard as distortion per se, but as soundstage shrinkage, loss of tone purity, and loss of depth. It is not necessary to boost the monitor gain and listen to low level material to perceive the degradation: the loss of depth in truncated material is audible at normal monitor gains. So you must learn the basics of managing wordlengths, and never turn your back on a computer! (Thanks, Bob Ludwig)

{ "Analog audio was easy, but digital audio is hard!" }

What about truncation at 24 bits? Within the same algorithm, truncation at 24-bits produces errors which are 48 dB lower than 16-bit errors. If 16-bit distortion products are barely audible, then wouldn't 24-bit errors be inaudible? On the contrary, I firmly believe that



Sample rate does not limit the timing resolution of digital audio as illustrated here where two waves are offset by less than half a sample period. Image courtesy of Dick Pierce ©2007.

> *"Digital audio's timing resolution is much finer than the human ear."*

truncation at the 24th bit *is* audible by trained listeners with a good monitoring system in a reasonably quiet room. I've noticed the losses of 24-bit truncation many times, without looking for them. Some Pro Tools users have complained about loss of stereo separation, but my listening tests indicate that a properly-dithered Pro Tools system has excellent stereo separation and depth. I believe this is completely within the user's power to control, if he studies Chapter 15. The more you know, the more power you have over your digital gear. Did you know there is a dither menu within the TC Electronic system 6000? As with all digital processors, you should look for the dither and wordlength menus. Looking deeper under the TC's hood, we know that Motorola processors are 48-bit double-precision fixed-point. So how does that 48-bit signal pass from engine to engine? Is the wordlength retained, or is it truncated to 24 bits? I've found that sending the signal from one engine to 24-bit dither, then out AES/EBU and back into the second engine sounds better than the standard method of routing the signal directly from one engine to the other, because the standard method truncates signal from 48 to 24 bits. I'm not completely paranoid about using DSP devices that truncate their internal signals to 24-bit, and some are sitting in my rack, but I do engage in a campaign with manufacturers to put 24-bit dither in their software, because it couldn't hurt, and it may just help!

Skeptics may want to undergo a blind test, so I invite you all to visit my demonstration called *Can You Hear Truncation?* which can be found at the links. Decide for yourself if you can hear 24-bit truncation.

Oversampling?

Processing via analog has one distinct advantage. Any nonlinear processor (including analog compressors) manufactures new high frequency energy content, but when this passes back via A/D conversion, the energy above Nyquist is filtered out by the analog anti-aliasing filter.[5] However, when processing digitally, this energy can fold back and cause aliasing distortion. The most advanced digital dynamics processors use oversampling technology, raising the internal sampling rate to reduce aliasing distortion. Jim Johnston says that even 8x oversampling is not enough to get rid of audible aliasing distortion. Current CPU power is not quite up to handling greater than 8x oversampling without sacrificing too many available plugin instances.

Why Is Good DSP So Expensive?

Intellectual property is the most nebulous thing to a consumer. It's easy to see why a two-ton Mercedes Benz costs so much, but the amount of intellectual work that has gone into a one-gram IC is not so obvious. It can take five man-years to produce a good digital equalizer; a good DAW has 500 or more man-years of development, created by individuals each with ten or more years of technical schooling or experience.

The Source-Quality Rule

An important corollary of this discussion is to underscore the importance of the *source-quality rule: Source recordings and masters should always have higher resolution than the eventual release medium. Start out with the highest resolution source and maintain that resolution for as long as possible into the processing.* When mastering, one consequence of this rule is to purposely reduce the number of processing generations, and if possible, go back one or more processing generations when a new process must be added or applied.

Ironically, the quality of the source counts the most when the end result is an inferior medium. For example, if you start with a high quality source (like a 24-bit mix) and dub to a lossy medium like mp3, it sounds obviously better than a copy from an inferior source (like an overprocessed 16-bit master). You'll never go wrong starting at 2496 even if the result is going to be a talking Barbie doll (of course there are diminishing returns).

VI. In Conclusion

The more you know, the better your work can be. But keep your perspective, internalize your technical knowledge and make it second nature, just like muscle memory — then you can work esthetically without a technical burden. Mastering engineers do not think about the meaning of life every time we go to work; we plug in our processors, listen, and make music sound better. But I also like to try to understand why things sound better, because it helps me avoid problems that are not obvious at first listen. Then I can dream up innovative solutions. I hope this chapter has inspired you to dream up some innovations of your own!

1
Moore's Law: The empirical observation made in 1965 by Gordon E. Moore, who postulated that the number of transistors on an integrated circuit for minimum component cost doubles every 24 months. Or a correlary, that computing power for the same cost and space doubles every 24 months (or even sooner!).

2
The ear's perception of noise is much more than just a frequency response curve, as Jim Johnston explains (in correspondence):

A single number is ineffective. Noise should be measured separately in each critical band and compared to the ear's threshold for that critical band.

3
Kraght, Paul (November 2000), Aliasing in Digital Clippers and Compressors. *Journal of the AES Volume 48 Number 11 pp. 1060-1065.*

4
Thanks again to Jim Johnston.

5
Thanks to Dan Lavry.

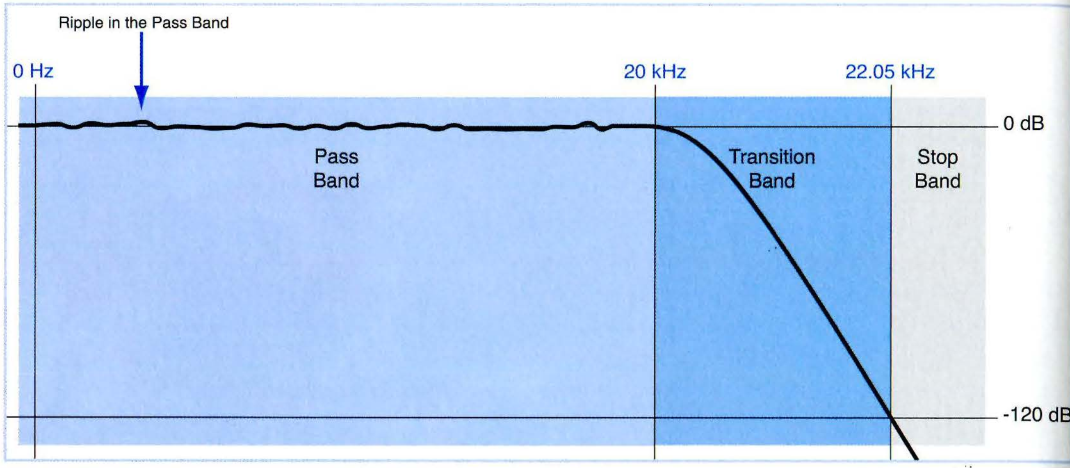
The Source Quality Rule:
"Always start out with the highest resolution source and maintain that resolution for as long as possible into the processing."

CHAPTER 23

High Sample Rates: Is This Where It's At?

I. Introduction

Regardless of the real benefits for the professional and the consumer, the current relentless drive for higher sample rates is lucrative for the hardware manufacturers. I've been working with higher sample rates for many years.[1] A great number of engineers (myself included) think that higher sample rate recordings sound better. Some of them point to the presence of supersonic frequencies in these recordings as evidence that we need the higher sample rates. Those of us who work with high sample rates cite the *open, warm, spacious, extended* sound of these recordings. But why is this so? How can our ears detect differences between 44.1 kHz, 96 kHz and even 192 kHz sample rates, since most of us can't hear above 15 kHz? Wouldn't it be nice if we could reconcile experienced engineer's preferences for high sample rates with the explicit knowledge that we cannot hear supersonic frequencies? This chapter presents some good explanations that are consistent with both points!



*Low-pass filter terminologies. The **passband** is the part of the frequency response which is not filtered or attenuated, in this example, from 0 Hz to about 20 kHz. This figure shows some **passband ripple** (non-flat passband frequency response). The **transition band** begins at the nominal cutoff frequency, until the **stop band**, where the response reaches the maximum loss of the filter. To avoid aliasing, the stop band must begin at or below the **Nyquist frequency**. The steepness of this transition band is the slope of the filter. In this example, the transition band is only about 2 kHz wide.*

I believe the answer to the dilemma lies in the design of digital **low-pass filters**, used in *oversampling* ADCs and DACs, illustrated in the above figure. Filters of lower quality, or which are unoptimized, exhibit tradeoffs such as low calculation resolution (which results in a smaller, colder sound), higher distortion, ripple, ringing, and potential for aliasing. The artifacts of ripple are time-smearing of the audio, and possible short (millisecond) echoes. Aliasing is a form of distortion that occurs if the filter does not have enough attenuation in the stop band (see sidebar on page 313). To avoid aliasing, we must use either a very steep filter, or a gentle filter with a higher cutoff frequency (which requires a higher sample rate). Many authorities feel that gentle filters sound better, because they have shorter impulse responses, and they claim that a long impulse response "smears" the signal in the time domain. However, I think anti-alias filter steepness is the least of our worries, within reason (see Bruno Putzeys' sidebar on page 314).

"If you hear an audible by-product of supersonic information, then something must be broken."

Ripple in the passband should be less than 0.1 dB.[2] It is harder to engineer a steep filter with low ripple, but it is perfectly doable; this can be achieved with a large number of filter taps. For the same number of taps, a more gentle filter will have less ripple. But more taps, lower ripple, and higher resolution require more components in the chip, which costs money, and to date converter chip manufacturers have barely stepped up to the plate. As we shall see, manufacturers who depend on these chips have used different workarounds to improve their lot.

Oversampling

One of the biggest improvements in digital audio technology came in the late 80s, with Bob Adams' **oversampling ADC**; this form of ADC has a front end which operates at 64 or 128 times the base sample rate. In other words, for 44.1 kHz operation, a 128X converter operates internally at 5.6448 MHz! The converter's noise is spread around a wider frequency spectrum and shaped, moving much of the noise above the audible frequency range. This high rate must then be digitally downsampled to the destination rate, at which time the supersonic noise is filtered out, to yield as much as 120 dB signal-to-noise ratio within a 20 kHz bandwidth.

Downsampling is done with a digital circuit called a **decimator**, a form of divider or sample-rate converter, which must contain an anti-aliasing filter. An oversampled DAC has an **anti-image filter** with an analogous role; though it operates at a higher sample rate, it too must have low distortion and ripple to sound good. While this was once costly to implement, the price of silicon is now infinitesimal compared to the benefits of good filtering. Still *the filters in chip converters made today are compromised.* There is no longer a reason for this practice – chipset manufacturers should begin doing it right.

To overcome the limitations of current converter chips, high-end converter manufacturers can either roll their own discrete converters (very expensive) or create workarounds using off-the-shelf components. Some manufacturers add DSP-based filters of their own design to supplement the chipset filters. For DACs, they upsample in front of the chip's own filter, so the chip filter does not have to work as hard. These hot-rodded DACs operate at 2x or greater rates regardless of the incoming rate. For ADCs, these manufacturers run the converter at higher rate, also easier on the built-in filter, followed by their own high-quality SRC. Though they help to make the total sound more transparent, supplementary filters are workarounds because the chipset filters are still in place.

An Upsampling Experience

Audiophiles, and some professionals, have been experimenting with digital upsampling boxes that are placed in front of DACs, supplying specious reasons to justify them. One argument by those who do not fully understand the nature of PCM is called "connect the dots." It goes like this: 'We need more dots than just 2 to properly describe a 20 kHz sine wave.' But this is erroneous: only 2 dots (samples) are necessary to describe an undistorted 20 kHz sine wave; when reproduced through a DAC, the low-pass filtering smooths out the waveform and eliminates all the diagonal lines.

In some cases, the proponents of an upsampling box report greatly improved sound. Although the improvement may be real, in my opinion it can be attributed to the various digital filter combinations, not to bandwidth or frequency response or (especially) to the sample rate itself. Remember that 44.1 kHz sample rate recordings, already being filtered, cannot contain information above 22.05 kHz. An upsampler cannot "manufacture" frequency information that wasn't there in the first place.

I've compared the sound of upsamplers against DACs working alone. Sometimes the upsampler makes an improvement, sometimes a degradation; sometimes the sound quality is the same either way. Sometimes the sound gets brighter despite a ruler-flat frequency response, which can probably be attributed to some distortion or to jitter between the first and second box. *In this digital audio world, sonic differences have come down to mathematics and clocking!*

II. Sample Rate Converters

There are two types of sample rate converters: *synchronous* and *asynchronous*. Synchronous models are the best-sounding types, but they cannot handle varispeeded rates. It is possible for synchronous converters to handle non-standard rates, such as pullup or pulldown rates required for NTSC video, but I have not seen any quality converters that can. Synchronous converters produce an exact multiple of output samples compared to the input, while asynchronous converters drift slightly over time as their filter coefficients vary so they can handle slight changes in incoming sample rate. Multiple passes through the same asynchronous algorithm may drift and phase slightly between each pass, but left/right channel pairs themselves usually retain their phase relationships, because filter sets

Nyquist, Sampling and Aliasing

Why filter on ADC

Sampling without filtering will include ALL signals, from the baseband that you want to keep, along with the out-of-band stuff you DON'T want, all the way out to infinity. This folds down (aliases) to the baseband, producing alias distortion, which sounds a lot like ring modulation, especially obvious on instruments like trumpets that have lots of high-frequency harmonics. That's why an antialias filter is needed when audio is sampled.

Why filter on DAC

The sampled audio stream which is played back contains the baseband and EVERY image of that baseband, all the way out to infinite frequency. That's why an anti-image filter is needed when going from sampled to continuous. Continuous means "analog."

The higher the sample rate, the higher the permitted filter cutoff frequency, 1/2 of the sample rate, known as the Nyquist Frequency.

The same basic rules apply to resampled digital streams. In other words, any sample-rate converter needs to properly apply anti-aliasing and anti-imaging filtering, because it involves re-sampling, very similar to the processes used in ADC and DAC.

Usually, the filtering is built into the resampler, and is not adjustable by the user.

Contributed by Dick Pierce

Bruno Putzeys on Anti-Alias Filter Design

I found that when you use fantastically steep anti-alias filters, you do get an audible sonic signature but it does not sound like smearing in any sense of the word. Instead, you get an overfocusing of the stereo image and a glassy top end. Some of the more naive FIR-based speaker DSP systems do sound like transients are smeared but this is clearly due to poor filter or cross-over design in the middle or bottom of the audio band, resulting in actual discrete (pre-) echos, not ringing at the extreme of the audio band.

I believe a 20 kHz LPF becomes transparent around the point when the transition band is 4kHz or more. This means that you can combine zero aliasing with a sufficiently short impulse response only for sampling rates of 48 kHz or more. For 44.1 kHz you have to allow some aliasing, which turns out to be less audible (on most material) than making the transition bandwidth 2kHz. All of the typical sonic

⟶page 315

are calculated in pairs and sometimes in multichannel groups. Still, to my ears, the stereo image of the asynchronous converter does not seem as stable. Multichannel asynchronous converter chips can synchronize the filters among channels on the same pass. In general I find their perceived depth is slightly less than that of a synchronous filter, but when upsampled to, for example, 384 kHz for use in a DAC, the errors of asynchronous converters become so small that DACs using this technique can sound excellent. You would have to spend 2 to 10 times the price of a chip-based DAC to get a small improvement in sound quality using a synchronous filtered model with discrete components. Both types of sample-rate converters can be either file-based (offline, non realtime) or standalone (realtime hardware). Most of the cheaper realtime hardware converters use asynchronous converter chips, which have come a long way in sound quality, but I think they should be avoided when possible. You will find realtime asynchronous converters embedded in many devices, e.g. audio interfaces which offer optional realtime sample rate conversion, digital consoles that permit mixing sample rates on different inputs, routers, and hidden within ADCs and DACs.

Asynchronous chips are used in routers common in video rooms, where rates have to be pulled down or up when making realtime transfers between recorders and devices. In these cases, crystal oscillators or wordclocks are used to lock the chip's output to a fixed rate. The table on page 315 summarizes converters, their uses, and my perception of their sound qualities.

III. Is It The Filtering or the Bandwidth?

Logic tells us that higher bandwidth cannot be the reason for superior sound at higher sample rates—since *the additional frequencies that are recordable by higher sample rates are inaudible.* It's fun and informative to put a high-pass filter into an orchestral passage that's got lots of cymbals and listen to the residual: it sounds like a shivery, swimming high frequency sound. But move that filter up past 10k, 15k, 20k (if you are very young). Do you still hear that residual? This is why I feel that people who stress the importance of seeing supersonic information in an FFT are barking up the wrong tree. One myth that's going around is the claim that, for example, inaudible 21 kHz and 22 kHz frequencies combine in air to produce an audible byproduct. But this is not true: air is linear at normal sound pressure levels—only nonlinearities (distortions) can cause an audible byproduct. In order to hear a beat frequency, both of the source frequencies must be audible and something must be broken! Either the power amplifier, the tweeter, or if it's a digital system, the filtering is inadequate and causing aliasing. If you hear beat notes, there is a nonlinearity. When testing, ensure that neither the amplifier nor the tweeter are producing distortion. Beat notes are not real frequencies, but rather a periodic variation in loudness whose rate is the difference between the two frequencies.

Here are some experimental confirmations from Bruno Putzeys:

> In actual fact, the beat note hypothesis has been tested using a supertweeter and a pair of sine-waves. The observations were: One amp, one tweeter: audible beating. Two amps, one tweeter: no audible beating. Two tweeters, each an amp: no audible beating. Conclusion: their amp produced audible intermodulation distortion, but the tweeter did not, and neither did

SAMPLE RATE CONVERTERS					
TYPE	FILE OR HARDWARE	RATES AVAILABLE (kHz)	REALTIME OR NON REALTIME	RECOMMENDED USE	SOUND QUALITY
Synchronous	File based	32, 44.1, 48, and their multiples	Non Realtime (offline)	Use this whenever possible	Superior quality is available. Quality varies among brands and models.
Synchronous	Hardware based (DSP or FPGA)	44.1, 48, 88.2, 96 and occasionally above.	Realtime	Preferable to Asynchronous when realtime is required if the application does not need a nonstandard rate. Used as upsampler inside highest-quality DACs and ADCs.	One standalone model (Weiss SFC-2). Superior quality, just below the best synchronous File-based model.
Asynchronous	File based or integrated in DAW as a real time pitch/time converter	32 through >192 including non-standard rates	Realtime and Non- Realtime are available	Use only when necessary to convert varispeeded or repitched material between non-standard rates.	Good to excellent
Asynchronous	Hardware (chip) based	32 through >192 including non-standard rates	Realtime	Use only when realtime conversion between DAWs or recorders is absolutely necessary (e.g., video applications, pulldowns and pullups such as 47.952 kHz). Also used in upsamplers in mid-to high-quality DACs and ADCs.	Good to excellent. The higher the destination rate, the less the loss of sound quality.

problems go away once you decide to do the filtering by the book, without the two main short-cuts typically taken in converter chips: halfband and equiripple. If you design both the decimation and upsampling filters using windowed sinc and begin the stop-band no later than fs/2, you can make a 48kHz AD/DA chain that's as good as undetectable.

BK: Since chipset manufacturers are generally not doing this, it may still be necessary to up the sample rate further to try to get the artifacts of these poor designs out of the range of hearing.

> *The issues of the audibility of bandwidth and the audibility of artifacts caused by limiting bandwidth must be treated separately. Blurring these issues can only lead to endless arguments."*
>
> — Bob Olhsson

Ultrasonic Audibility?
David Griesinger demonstrates (see links) how easy it is to reach the wrong conclusion about ultrasonics and how hard it is to devise a test which would confirm that ultrasonics (or their byproducts) are audible. He cites the distortion of loudspeakers in slides 3-6 of his presentation. Essentially, he found the same issues as Bruno, that ultrasonic harmonics were audible when played through the same loudspeaker as the low frequency tone, but not audible when played through a separate loudspeaker. In other words, the ultrasonic harmonics were only audible when played through a system where non-linearities caused artifacts in the low frequency range; when presented directly to the ear, they were not audible. The rest of Griesinger's presentation is also very informative.

> the ear. Mixing supersonic signals to get audible content has been done, but this technique relies on the nonlinearity of air and in order to make the air nonlinear, the SPL involved is enormous (over 140). The quality isn't good but then again the target application was crowd control.

There have been other arguments about the importance of supersonic signals. But why try to invent reasons when there are scientific and logical explanations for why high sample rates sound better that make perfect sense and have even been proved by experiment! One experiment I discuss below shows that even recordings without supersonic information can sound better when reproduced at the higher sample rate!

In December 1996, I sought to systematically find reasons for sonic differences between sample rates, performing a listening test, with the collaboration of members of the Pro Audio maillist. The question we wanted to answer was: *Does high sample rate audio sound better (or different) because of increased bandwidth, or because of less-intrusive filtering?* We developed a test that would eliminate all variables except bandwidth. Other major factors were held constant: sample rate, filter design, DAC, and jitter.

The test we devised was to take a 96 kHz recording, and compare the effect on it of two different low-pass filters. The volunteer design team consisted of Ernst Parth (filter code), Matthew Xavier Mora (shell), Rusty Scott (filter design), and Bob Katz (coordinator and beta tester). We created a digital audio filtering program with two impeccably-designed filters that were mathematically identical, except that one cuts off at 20 kHz and the other at 40 kHz. The filters were designed for overkill, with exemplary characteristics: double-precision dithered to 24-bits, FIR linear phase, 255-tap, >110 db stopband attenuation, 2 kHz transition band, and <.01 dB passband ripple.

For the first listening test, I took a 96 kHz sampled orchestral recording, filtered it and laid both versions into a Sonic Solutions DAW for comparison. I expected to hear radical differences between the 20 kHz and 40 kHz filtered material. But I could not hear any difference! Next, I compared the 20 kHz filtered against "no filter" (of course, the material has already passed through two 48 kHz filters in the ADC and the DAC). Again, I could not hear a difference! The intention was to listen double-blind; but even sighted, 10 additional listeners who took part in the tests (one at a time) heard no difference between the 20 kHz digital filter and no filter. And if no one can hear a difference sighted, why proceed to a blind test?

I then tried different types of musical material, including a close-miked recording of castanets (which have considerable ultrasonic information), but there was still no audible difference. I then created a test which put 20 kHz filtered material into one channel of my Stax electrostatic headphones, and the time-aligned wide-bandwidth material into the other channel. I was not able to detect any image shift — there was always a perfect mono center at all frequencies in the headphones! This must be an amazing filter!

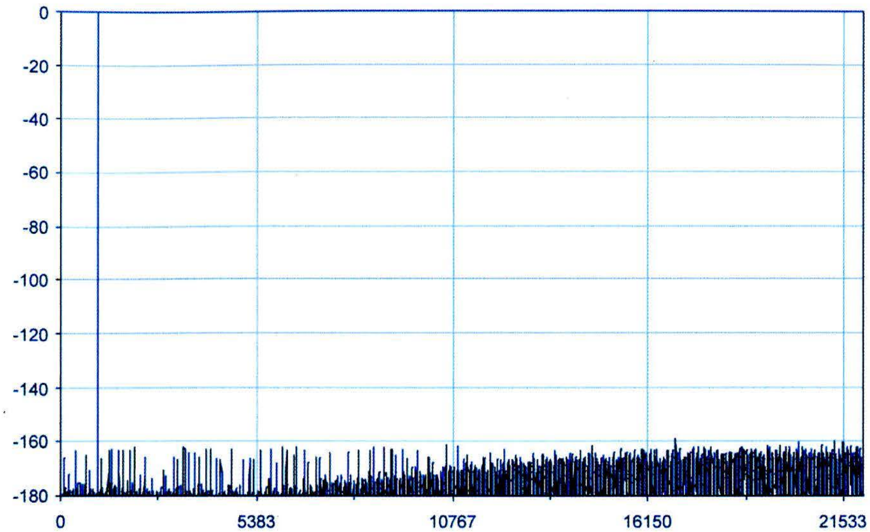
As a last resort, I went back to the list and asked maillist participant Robert Bristow Johnston to design a special "dirty" filter with 0.5 dB ripple in the passband. Finally, with this filter, I was able to hear a difference: it added a boxy, veiled, "gritty" quality that resembles the sound of some of the cheaper CD players we all know.

After I conducted my test, several others tried this filtering program, and most reached the same conclusion: *a well-designed filter is inaudible.* One maillist participant, Eelco Grimm, a Netherlands-based writer and engineer, performed the test and reported no audible differences using a Sonic Solutions system, yet he and a colleague passed a blind test between filtered and non-filtered using an Augan workstation. He did not compare the sound of the 20 kHz versus 40 kHz filters, so we are not sure if he was hearing the filter or the bandwidth (I suspect the filter). We are not certain, but perhaps the reason Eelco uniquely reported a sonic difference is that the Sonic system produced sufficient jitter to mask the other differences, which must have been very subtle indeed! Be aware that the filters in the converters themselves are chained with the filters under test, and so may have obscured the audible effect of the test filter — so it is very difficult to design a single variable listening test.

This 1996 test seems to show that a "perfect 20 kHz filter" can be designed. We are just not sure how good the filter has to be to be considered inaudible. Regardless of whether Eelco's group did reliably hear differences, it should be clear by now that differences people hear between sample rates are more likely due to filter design than to supersonic bandwidth. Ironically, it was necessary to make a high sample rate recording to prove that high sample rates may not be necessary!

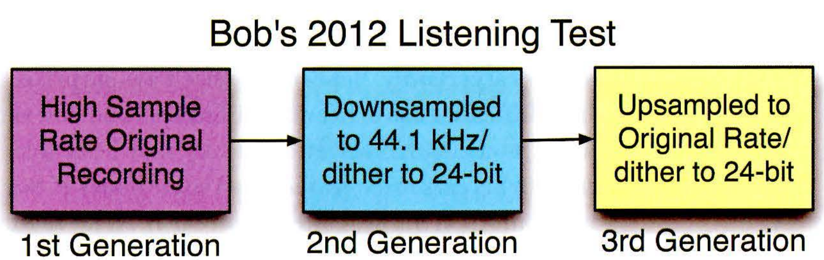
Despite this evidence, this issue remains controversial after more than 16 years. Many respected engineers still believe that supersonic information is important to a recording. So in 2012 I devised a modern version of the test, using what I consider to be the world's best sample rate converter as a filter, the Weiss Saracon, illustrated above right. All other variables were kept

1kHz Sine 0dB Converted From 96kHz to 44.1kHz



The textbook-perfect distortion and noise performance of a Weiss sample frequency converter. With this SFC it is possible to convert between non-integer rates with identical measured performance. In other words, no difference between downsampling from 88.2 kHz or 96 kHz to 44.1 kHz.

constant. I began with high sample rate recordings (at 88.2, 96, 176.4 and 192 kHz), many of them showing information above 20 kHz on an FFT. I call these files the *1st generation*. I then downsampled this material to 44.1 kHz/dithered to 24-bit with Saracon, the *2nd generation*. This removes all information above 22.05 kHz using an exceptional-quality low-pass filter. I then took the 2nd generation file and upsampled it back to the original rate; this is the *3rd generation*. I then inserted the 3rd and 1st generation files in a Sequoia DAW, matched in timing. A null test, with an FFT measurement of the difference, revealed just a supersonic remainder with nothing but dither noise in the audible band! We then



MYTH: *Upsampling makes audio sound better by creating more points between the samples, so the waveform will be less jagged.*

switched between 1st and 3rd generation, single blind. I (and several others) were quite surprised that we failed to hear a difference between the 1st generation (source) and the 3rd generation (result), playing at the original higher sample rate.

I then played the 2nd generation (44.1 kHz) file, in another instance of Sequoia. *This file sounded worse to my ears than 1st or 3rd generation!* Wait a minute — how can a file sound worse than the result derived from it? Now we're getting somewhere: We can only conclude that *low sample rates sound worse because of how they are reproduced* on a typical chip-based converter. I also compared the sound of the 2nd generation file on various chip-based converters, and to my ears, the converter having the best and highest upsampling filter sounded the best, the most like the 1st generation, though not quite identical. Again, *the evidence points in the direction that the problem with low sample rates is in the reproduction, not the rates themselves*. I did not have available a hyper-expensive discrete converter with "roll-your-own" filters, but this would have added more valuable data to the listening tests. It would also be interesting to try a 48 kHz second generation. I'll leave those possibilities to another experimenter.

I passed this information on to researcher and designer Bruno Putzeys, who subsequently spent several months performing the same listening test, double blind, with a long-term listening method. After months of listening, Bruno was able to hear a difference between the 1st and 3rd generation result with some statistical certainty. But Bruno himself was surprised by how difficult it was to hear the difference, and how very small it was. With much effort he may have been able to train his ears to detect the small sonic losses of two generations of the best filtering on earth. As far as I'm concerned, these two tests settle the issue: supersonics are probably not important, but filtering quality definitely is! I urge anyone who doubts the results of these tests to try them yourselves; the tools are readily available. I can send sample files to individuals who do not have Saracon and wish to perform the listening tests.

IV. Psychoacoustics, Other Filter Solutions

Audio researcher Jim Johnston,[3] who knows as much about the time-domain response of the ear as anyone, has shown that steep low-pass filters at or near the high frequency limit of the ear interact with the cochlear filter, creating pre-echoes that the ear interprets as a loss of transient response, obscuring the sharpness or clarity of the sound. Jim has experimentally calculated that the minimum sample rate which would support a Nyquist filter gentle enough to elude the ear is 50 kHz. If he is right, then the 48 kHz professional rate is nearly sufficient. Jim based his conclusion on the length of the shortest organic filter in the human ear, and notes that the 50 kHz number nicely matches the original work with anti-aliasing filters done by Tom Stockham for the Soundstream project.

We also have to consider cumulative effects, for even if an inaudible filter can be designed, will 2, 3, or 4 in series also be inaudible? Perhaps this is irrelevant, as researcher Peter Craven has discovered: *Ringing or pre-echo problems in a filtered system can be completely eliminated by adding a properly-specified gentle slope filter anywhere in the record or reproduction chain.*[4] This seems counterintuitive, but Dr. Craven has the mathematical proficiency to prove this, so his paper ought to have a profound effect on how converters and digital systems are designed. His discovery may explain why some digital audio systems sound better than others; it may explain the discrepancy between my listening test and Eelco's. For example, if Eelco was listening through a DAC with a steep filter, and I was listening with a gentle one — that could override the effect of the sharp filter under test. Craven's discovery alone is justification for using a 96 kHz sample rate (in order to allow a sufficiently gentle filter at the tail of the chain), or upsampling to that rate for reproduction. Converter and systems manufacturers must review this research and provide tools for better-sounding digital audio; all it takes is the impetus to do it right.

How good is 44.1 kHz sample rate? The answer: It's gotten a lot better with improved upsampling DACs, I've found the audible difference between a 96 kHz original and a 44.1 k result has become subtler. Once again this points to the filters as the culprits, not the sample rate. Therefore, a well-designed DAC should exhibit very little audible difference between sample rates. Can 44.1 kHz ever sound equal to 96 kHz? Probably not, but 48 kHz may eventually sound equivalent (as Bruno pointed out in his sidebar).

What is the best rate for practical engineers to use today? Even recording and mixing engineers working on non-audiophile projects should upgrade to at least 48 kHz from their customary 44.1 kHz. This avoids the sharpest filters, also gives a bit more "professional room," and probably will result in better sound after several generations of processing and/or mastering. For audiophile and jazz/classical projects I recommend at least 96 kHz sampling. When it comes to mastering, I upsample mixes to 96 kHz to reduce further degradation.

Is 192 kHz sampling necessary? Many engineers are quite enthused about that rate. Keep in mind that all modern converters are upsamplers, so the only difference between a manufacturer's 192 kHz and 96 kHz DAC may be the decimator! Dan Lavry has found that distortion increases and conversion accuracy decreases unless a chip is used which has been optimized for a particular rate.[5] So if you decide to record at 192 k, ensure that your converter performs better at this rate than at 96.

The Advantages of Remastering 44.1 Recordings at Higher Rates

Researchers such as J. Andrew Moorer of Sonic Solutions, and Mike Story of dCS have demonstrated theoretical improvements from working at a higher sampling rate. Moorer pointed out that post-production processing, such as filtering, equalization, and compression, will result in less distortion in the audible band, because the errors are spread over a higher bandwidth — and at least half of that bandwidth is above 20 kHz.[6] In addition, if the destination after processing is a high-resolution medium, then the master can be left at the higher sample rate and wordlength, avoiding another generation of potentially sound-veiling 16-bit dither and the consequences of low-pass filtering at the end of the process. Thus, consumers should not scoff at masters that have been digitally remastered from original 16-bit/44.1 kHz sources. They will be getting real, audiophile-quality sonic value in these remasters, even if they do not see supersonic products in their FFTs.

"The filters in a typical compact disc player or in the converter chips used in most of today's gear are mathematically compromised."

1
I was the recording engineer for the world's first 96 kHz/24-bit audio-only DVD.

2
According to Julian Dunn. Some say that less than 0.01 dB is a requirement for ripple to avoid sonic artifacts.

3
In correspondence. JJ is the inventor of the science of perceptual coding, which led to coding developments such as mp3, AAC, etc.

4
Craven, Peter (2003) Controlled Pre-response Antialias Filters for Use at 96 kHz and 192kHz, AES 114th Convention Preprint 5822.

(2004) Antialias Filters and System Transient Response at High Sample Rates, JAES Volume 52 Number 3 pp. 216-242

In correspondence: Dear Bob, yes I am the person responsible for apodising - which is an extremely simple idea. Given a brickwall filter, if you taper down the response before the brickwall, there will be less audio energy exciting the brickwall and producing ringing. You can also skew the phase response before you get to the band-edge so as to reduce pre-ringing, which seems to be perceptually more destructive than a post-ring. This is somewhat similar to using a gentle filter rather than a sharp one, but there is an important new idea. In a context where you might have to take the sample rate up and down several times, if you cascade several gentle filters you are in danger of ending up with something sharp. Whereas a brickwall cascaded with a brickwall is still a brickwall. So the prescription to handle this case is to use brickwall filters as close as possible to Nyquist throughout the chain, but to have a single apodising filter somewhere to take the energy gently down to zero. It matters not whether the apodising filter is before or after the brickwall filter(s), but ideally that should be standardised so that we don't get two of them. I have suggested that it should be done at the playback end, though that is arguable. My papers deal explicitly with the case of taming the ~48kHz brickwall that is implied by 96kHz sampling.

5
A link to Lavry's papers can be found in the links. Lavry also warns (in correspondence):

> Using an ADC designed for 192 kHz operation at 96 kHz is not the same as using an ADC intended to operate at 96 kHz! The 192 kHz design already has the accuracy tradeoffs; using it at 96 kHz does not remove the tradeoffs, it is in fact the same conversion with a X2 additional decimation.

> ...anyone that decimates or does sample rate conversion from 192 kHz to 96 kHz or 48 kHz or 44.1kHz [who] says they hear a particular sound from that original 192 kHz conversion, is in fact supporting the fact that what they hear resided all along under the new Nyquist.

A related reaction comes from Crispin Herrod-Taylor (in correspondence):

> Probably the only real way of proving whether 192 K sounds better than 96 K is to get a good ADC chip [with no compromise] running at 192K, and then add a DSP which has optimised filtering for the 192K and 96K.

6
Julian Dunn (in correspondence) clarifies:

> A 3 dB reduction in distortion results because the error products are spread amongst twice the bandwidth. This is true for **uncorrelated** quantization errors which fall evenly throughout the frequency range from dc to fs/2. And does not work for distortion products which will correlate with the signal [such as from compressors].

Nika Aldrich (in correspondence) qualifies:

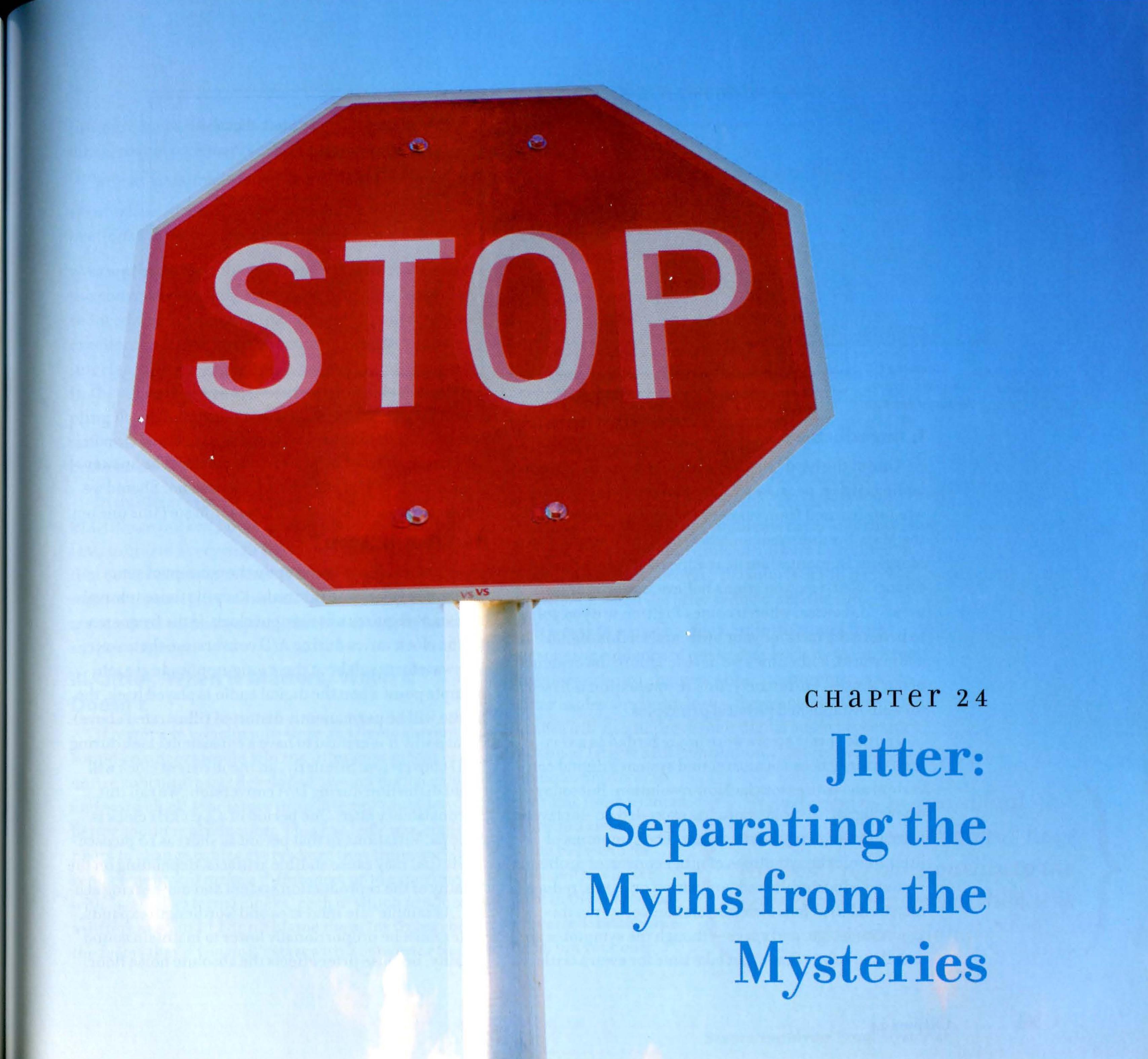
> ...increasing the sampling frequency [simply] in order to increase dynamic range is an exercise in futility: the effect is swamped by other forms of noise anyway.

Jim Johnston (in correspondence) indicates:

> processing at higher rates is *required* for any non-linear processing, such as compression. These non-linear processes produce new frequency components, some at higher frequencies.

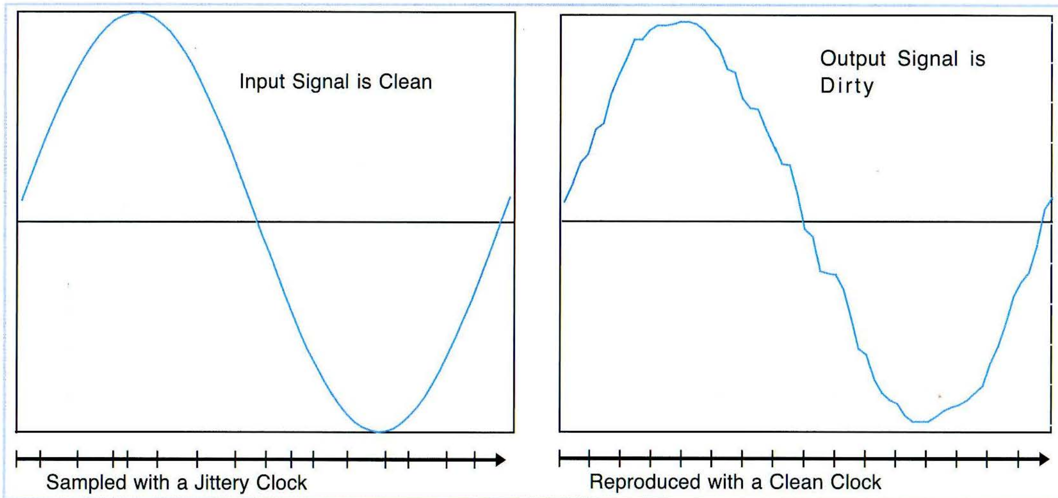
Conclusion: Other arguments aside, a high enough sampling rate for processing is required to avoid aliasing of these new frequency components.

"A well-designed DAC should exhibit very little audible difference between sample rates."



CHAPTER 24

Jitter: Separating the Myths from the Mysteries



Jitter during A/D conversion creates permanent distortion

I. Introduction

One of the hardest-to-explain phenomena in digital audio is **jitter**, because we have to reconsider much of what we learned from years of analog experience. In the Marx Brothers movie *Duck Soup*, Chico confuses Margaret Dumont when he reappears to her disbelief, saying "Well, who you gonna believe, me or your own eyes?"[1] Likewise, when it comes to jitter, are you going to believe the facts, or your own ears! In this digital audio world, sometimes we have to ignore the evidence of our senses. Fortunately, this re-evaluation is based on well-established physical principles.

In 1980, jitter errors were not regarded as a very high priority because most sound system's digital converters and processors had low resolution. But today, where signal-to-noise ratio has exceeded 20-bit level, jitter problems are more evident. The symptoms of jitter mimic the symptoms of other converter problems — sound which is blurred, unfocused or harsh, reduced image stability, loss of depth, ambience, stereo image, soundstage, and space — though the symptoms are usually so subtle that it can take time for even a critical ear to learn to identify them. Is our digital audio actually being affected by jitter in our clocks? The answer is: sometimes yes, but most of the time no! Should we believe our ears? It'll take a Chapter to sort this one out.

II. What is Jitter?

Digital audio is based upon the concept of sampling at regular time intervals. Keeping those intervals constant requires a consistent clock. If the frequency of the clock varies during A/D conversion, then since the waveform will be at the wrong amplitude at each sample point when the digital audio is played back, the audio will be permanently distorted (illustrated above). That's why it is critical to have a consistent clock during A/D conversion. Similarly, an inconsistent clock will yield distortion during D/A conversion. We call this inconsistency *jitter*. One period of a 44.1 kHz clock is 22.7 µs. Variations in that period as short as 10 picoseconds (ps) may cause audible artifacts, depending on the quality of the reproduction system and our hearing acuity. As sample rate increases and wordlength expands, jitter must be proportionally lower to maintain sound quality, because jitter affects the absolute noise floor.

Jitter produces sidebands (additional frequencies, or tones) that mask inner detail in a recording.

We can measure jitter in two places:

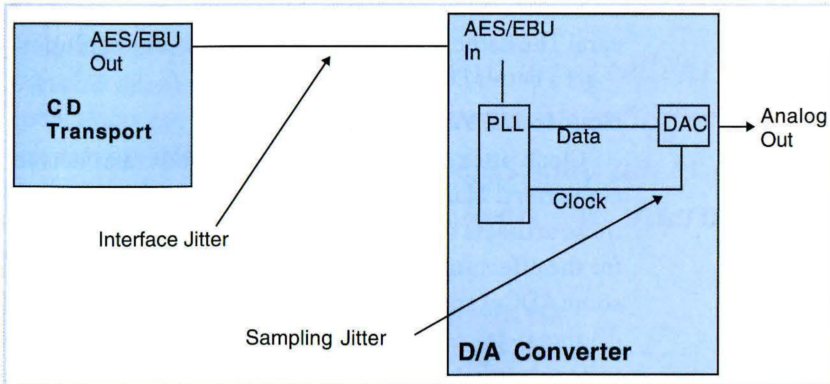
1) **interface jitter**, the jitter present in the interconnections between equipment.

2) **sampling jitter**, the jitter in the clock that drives the converter. Luckily, sampling jitter can be so reduced that it becomes inaudible: if a converter has excellent internal jitter rejection, then even high interface jitter may not cause audible sampling jitter. In this Chapter, we are mostly concerned with sampling jitter, because interface jitter is rarely important unless it causes a breakdown in communication between devices.

In this figure (pictured upper right) it is up to the PLL (Phase Locked Loop, to be explained) inside the DAC to create a very high frequency sampling clock that drives its components. If it is a superb PLL (rare), none of the artifacts of incoming interface jitter will be transmitted to the sampling clock or cause distortion in the converter.

III. Jitter, When it Matters, When it Doesn't

If leaping to conclusions were an Olympic event, sound engineers would win the jitter gold medal — an entire subculture has developed around digital cables and word clock generators in an attempt to achieve better sound reproduction. This has led some engineers to change cables everywhere they think such a replacement will make a difference, or to experiment with "stable" external clocks, each of which produces a different sound.[2] I don't blame them for trying, but the fact is that if you don't start with a converter that has a good PLL, no external box can help. Many of these clock generators hurt the sound if you pick the wrong unit and/or it is not compatible with the PLL in your converter.



Interface jitter vs. sampling jitter

Within the AES/EBU or S/PDIF interface, the embedded clock interacts with the data stream.[3] No cable can remove the jitter problems in the interface, so external jitter reduction units will always be limited in their effectiveness, because the interface will increase the jitter between the jitter reducer and the DAC.

It's easy to see where the next myth comes from, because the evidence of our senses leads us to the wrong conclusion. Since engineers hear improvements with better cables and wordclocks, they conclude these devices will also improve their digital audio processors. But this is largely a misconception. *Audio processors process data, not clock*, so any sonic improvement is due to a cleaner clock being passed to the DAC, not to a difference in the data being processed. Believe the facts, not your own

> *"Traditional audio signal-to-noise ratio measurements have (almost) no relationship to the sound of a converter when it is receiving signal."*

ears! The listening problem has an immediate solution — get a better DAC!

How to Lie With Measurements

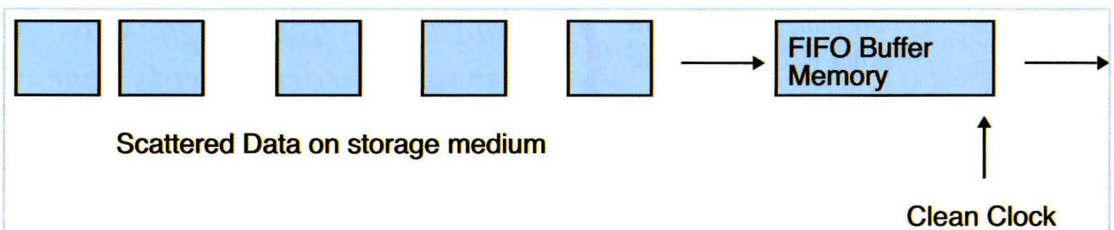
Clock jitter can produce insidious audio artifacts in converters. Manufacturer's specifications often hide these artifacts because there is no established criterion for the effects of jitter on converters. For example, some ADCs (and a few DACs) report exceptional >120 dB signal-to-noise ratios, theoretically equivalent to >20-bit performance. But is this true in practice? These figures are obtained by the traditional method of calculating signal-to-noise ratios: first measure a full-scale sine wave signal, then remove the signal and measure the residual analog noise. But this method does not take into account *the noise modulation and distortion when a clock is jittery and the audio signal is complex (such as music)*, which accounts for some of the previously-unexplained sonic differences between converters that measure identically. Most *signal-to-noise ratio* measurements quoted in manuals are therefore irrelevant, and most people have never heard true 20-bit performance, let alone 24. *Signal-dependent jitter* is the worst-sounding type of jitter, producing a blurred quality to the sound.

A converter that successfully rejects jitter can sound much cleaner than another with a lower static noise floor. Jitter can produce *random effects* (which translates to a higher random noise floor that can also be signal-dependent), and *discrete frequency effects* (such as other clocks in the box producing random tones and inter-modulation between the other clocks and the main sampling clock). Some of these effects are more benign to the ear than others, which is why it is so difficult to put a single number on jitter performance.

Storage Media

There is no jitter on a storage medium — only the data is stored, not the clock (there is no clock on a compact disc).[4] An entirely new clock drives the data when it is played back from the storage medium. The manner in which data is stored on the medium has no effect on the output jitter. Bits are usually stored in a very irregular fashion: on hard disks, the data may be out of order, non-contiguous, and widely spread. Data stored on CD (in EFM format) must be unscrambled and decoded during playback, but scattered storage has nothing to do with *jitter*, since time is not involved until the data is played back.

During playback, widely scattered data is collected into a buffer memory whose output is controlled by a steady clock (pictured below). The quality of that clock and its driver circuitry is the origin of any interface jitter. Audio manufacturers differ widely in their abilities to keep outgoing clocks under control, but all face the obstacle that clock stability is not important to the computer-based technology we have adapted to digital audio. In fact, the standard computer hard disc interfaces are *asynchronous* (non-clocked), having a completely irregular output with enormous equivalent jitter. When such non-clocked interfaces



Spacing of data on storage medium has no meaning as it is first buffered and output to the world with a clean clock.

are used, it is the task of following circuitry to make the data conform to a steady clock.

Although there is no jitter on a storage medium (since it has no clock), if we want to isolate the causes of playback jitter, we have to study the complete mechanism. We may discover some interference pathways within a player where storage issues could, indirectly and minutely, affect the output clock because of poor power supply or grounding design; this may potentially cause audible artifacts as a tertiary effect. But as we shall see, a well-designed digital audio receiver should be able to reject jitter at such low levels. Caveat designer!

Clock Stability Requirements for Converters

An ordinary crystal oscillator is sufficient for running a computer that processes data, but audio converters require an extraordinarily stable master oscillator. To get 20-bit performance at 44.1 kHz requires oscillator stability (jitter) at or below 25 ps peak-to-peak.[5] One nanosecond (1000 ps) in the time domain equates to 1 GHz, which is why a critical converter's circuitry must be shielded and isolated from even the tiniest RFI or clock leakage that can enter via power supply, grounds, or emissions. It should now be obvious why good-sounding converters are rare and expensive, and why the converters on most computer-cards do not sound very good: there are a lot of interference and power supply issues within a computer chassis.

IV. How to Get the Best Performance from Converters

There are two ways to clock a converter:

a) **Internal Sync**, where a (hopefully) stable crystal clock located inside the converter (very close to the sampling clock pin of the converter chip for the best audio performance) directly drives the circuitry. This is not very costly in parts but does require good layout, grounding, and power supply design.

"Only in an excellent converter design can jitter performance via PLL be as good as, or negligibly worse than, via internal clock."

b) **External Sync**, which as we have seen cannot be used directly, and requires a PLL, perhaps the fundamental culprit of jitter-induced converter artifacts. The PLL has to filter jitter caused by poor source clocks, by the AES/EBU line itself, or by interference along the cable that brings in the clock. The common use of unbalanced wordclock cables can produce ground loops in the clock signal itself, unless the converter can properly reject the hum products.

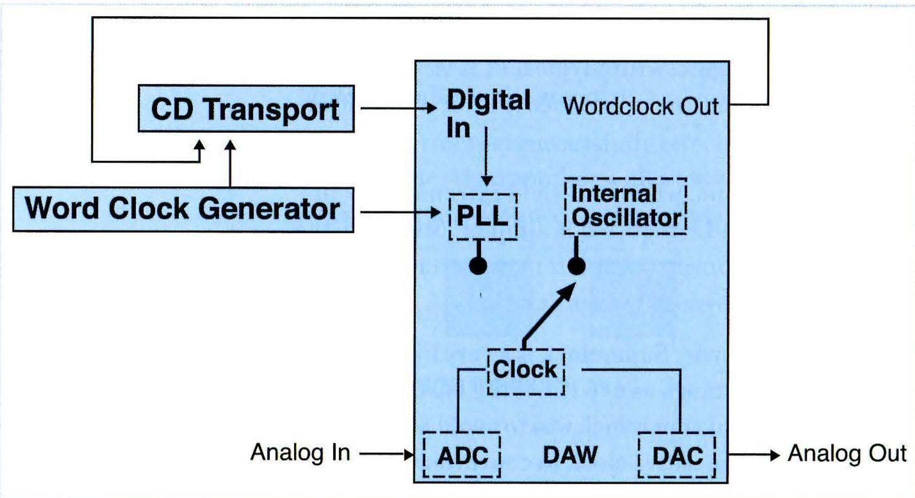
Examples of External Sync:

i) **AES/EBU sync**, which is prone to *signal-related jitter*, also known as *program-modulated jitter* or *data-dependent jitter* as first illustrated by Chris Dunn and Malcolm Hawksford in their seminal paper.[3] Thus AES/EBU "black" will produce a cleaner clock than AES/EBU with signal, with a typical PLL. A "smart" PLL in a converter can still reduce this interface jitter to inaudibility.

ii) **Wordclock sync**, which can yield extremely low jitter, because the PLL required is simpler. Still, only the best-designed converters are immune to audible jitter when externally synced via wordclock.

iii) **Superclock sync**. Superclock is a very high frequency clock, as much as 256 times the base sample rate. The object of superclock was to avoid a PLL entirely, since the native clock in converters is already at this high rate. However, there is no such thing as a free lunch — superclock is very fragile, and manufacturers must still pay attention to jitter issues with superclock.

iv) **Other Interfaces.** The embedded clock in Firewire, USB, or any other interface should not be used to drive a system. For lower jitter, a separate wordclock cable is needed or the converter should be on internal sync. Make sure Firewire carries only the data, but not the clock. The most meticulous manufacturers of interfaces carefully separate data and clock issues. In the past few years, audiophile-quality USB DACs have become popular, and if these use an **asynchronous** USB connection, jitter can be extremely low. In asynchronous USB, just as with isochronous Firewire, the converter runs on internal sync or wordclock sync with a clean clock driving the converter chip. When it needs more data, it sends asynchronous data request signals back to the DAW or media player. A PLL is not required in this case. Thus, asynchronous USB and Firewire can potentially sound better than a converter connected by AES/EBU or SPDIF, but it's getting harder to hear the difference because PLLs have also greatly improved.



DAW with ADC and DAC in a single box

Single Box Solutions

In this diagram of a DAW (pictured bottom left), the ADC and DAC are enclosed in a single box, so one clock drives them both (this applies to 2-channel or multichannel converters).

To record analog in, the DAW clock may be set to:

- internal oscillator (a.k.a. internal sync or clock)
- external sync via wordclock
- external sync via AES/EBU or SPDIF

In most cases, the cleanest-sounding option for analog recording is to use internal clock — unless you are certain that the external clock is superior and the PLL in your DAW's interface is the type that will benefit from external clock.

To record from a digital source, the DAW clock must be derived from the source. For example, from a CD or DVD transport, the DAW clock may be set to:

- internal oscillator, in which case the CD transport must receive external sync from the DAW's wordclock output. Since very few CD transports have wordclock inputs, this option is usually not possible.
- external sync via wordclock, in which case the CD transport may receive external sync from the DAW or the wordclock generator. Again, this option is usually not possible.
- external sync via AES/EBU or SPDIF: for example, the CD transport becomes the source of sync. This is the most common option for dubbing in real time.

Jitter on the interface has no effect on any transfer. Only the data is transferred, not the clock. This concept is getting hard to communicate because realtime transfers have become passé, since discs can be imported into the computer in non-realtime. But DAW to DAW transfers and transfers to and from samplers are still made in real time, so you do need to learn how to configure interfaces and their clocking. Start with this principle: There must be only one master clock; every other device must slave to that clock.

Multi Box Solutions

Jitter is harder to optimize with ADC and DAC in separate boxes, pictured at right.

For recording from analog sources, the clocking options are:

- ADC as master on internal oscillator, with DAC slaving via wordclock or directly by digital input
- Word Clock Generator driving all boxes

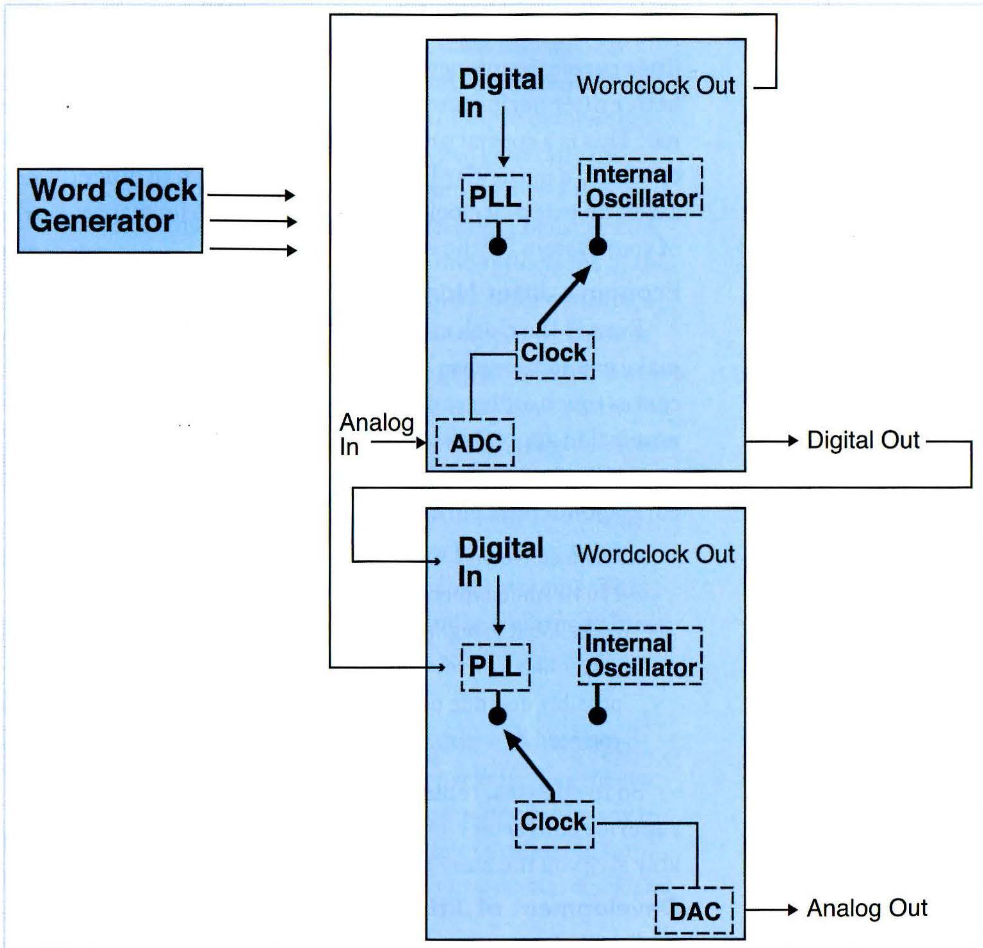
In practice, the cleanest sound is usually produced with the ADC as master clock for recording and the DAC as master clock for playback. But standalone DACs (except for USB DACs) normally do not have an internal sync option, so the simplest playback solution is to use a jitter-immune DAC locked via its digital input.[6]

Atomic Nonsense

One of the greatest myths perpetrated by manufacturers is the need for an "atomic clock": an external clock whose accuracy is very well specified. If its output is supposed to be 44.1 kHz, then they may specify it as 44.100 or whatever accuracy. But clock *accuracy* is insignificant for mastering purposes; what counts is the clock's *stability*—its jitter (short-term variances in frequency). Ironically, a clock can have high jitter but still be considered accurate.[14]

Performance of the PLL with External Sources

There is no guarantee that your equipment's performance will improve when locking your equipment to an external clock. In many cases it can degrade, depending on the nature of the clock and the PLL (see links). Since it is far more difficult and expensive to build a good PLL than a stable crystal oscillator, *only an excellent converter design can deliver jitter performance via PLL as good as via*



Jitter is more difficult to control when ADC and DAC are in separate boxes

"Digital processors look at the samples, not their time of arrival."

internal clock. Designer Bruno Putzeys has shown that in one unique case it is possible to slave a PLL and get better jitter performance at *low frequencies below those at which the PLL operates* (called the *corner frequency*). In this case, the low-frequency jitter of the source controls that of the receiver. So if the external clock is a superb oscillator with very little low frequency jitter, and the receiver is a typical DAC with a medium jitter corner frequency, then the receiver might exhibit better jitter performance on external sync than internal! This is a special case, only with a particular external clock and a particular DAC. Keep in mind that in other cases, an external clock could change the performance of your system for the worse.

Economic Jitter Nonsense

Even if the clock makes the sound better, does this make economic sense? I assure you that the clock could cost as much as the converter. It would be far more economical to apply the cost of the external clock to a better converter running on internal sync. Eelco Grimm (in correspondence) points out...

> Prism converters have a corner frequency below 200 Hz while typical converters' PLLs are above 2 kHz! So it is highly likely that a very good converter like a Prism will not be affected at all, or possibly degrade no matter what external clock you feed it.

So in all cases, replacing an existing converter with a superior converter running on internal sync will probably give you the most bang for the buck.

Development of Jitter-Immune Converters

DACs. Advances in design have produced affordable DACs with good-sounding analog circuitry and virtual immunity to incoming jitter. In addition to their traditional analog-style PLL, these DACs have a stable internal crystal clock and a digital ASRC (Asynchronous Sample Rate Converter) for secondary PLL and anti-imaging filter. Although ASRCs add some distortion and image wander, they have greatly improved. This style of DAC has all but displaced any cheaper DACs with traditional "purist" PLLs because only a few skilled designers know how to make a good-performing traditional PLL. Making a quiet, jitter-immune traditional PLL requires good analog design skills, including power, grounding and circuit board layout. This has not daunted the few remaining purist manufacturers who make traditional DACs, most of which are very expensive. Probably the most economical full featured "purist" interface is the Prism Lyra series, whose sonic performance demonstrates to me that a high-quality traditional PLL sounds better than the best ASRC-equipped DAC, with a wider, deeper and more stable stereo image. It's the first DAC/ADC combo that sounds audibly transparent to me at 96 kHz — it's very difficult to tell if it is in the circuit.

ADCs. The state of the art in ADCs is also undergoing a revolution with some manufacturers experimenting with a topology that uses a crystal oscillator to drive the converter chip at a high sample rate for lowest distortion and best performance. The converter chip is then connected to an ASRC chip to synchronize the data to an external clock.[7] The ADC data is sample-rate converted, and the ASRC functions as a downsampling filter. The jury is out on the sound quality of this topology, but preliminary impressions are that at least it sounds good. Keep in mind that compared to traditional-style ADCs with synchronous filters, ASRCs change the data in a radical way. If this technology becomes established, then good-sounding, inexpensive ADCs which can lock to external clock will be easily available. But can they sound excellent considering the quality of current ASRCs? For the rest of this chapter I will be referring to "traditional-style" ADCs (which do not incorporate an ASRC).

To make a superior (traditional) ADC that produces only *inaudible* jitter on external clock requires time, research, and critical design implementation of PCB layout, grounding, internal clock distribution, and rigorous separation of digital and analog signals. The engineers who produced a superior converter model I'm familiar with spent one *man-year* on the phase locked loop alone, and a further year on the converter details.

Digital Cables

How important are digital cables? Mismatched impedances can cause signal reflections in a cable that can result in jitter or, in some cases, poor data transfer. The higher the sample rate and/or the longer the cable run, the more critical it is to have properly matched impedances. So if you can get an audibly-clean data transfer at, say, 192 kHz sampling, then this is probably a good digital connection! More formally, technicians can check the **eye pattern** to ensure proper electrical signal quality. Of course it's important to use the correct impedance cable and connector; however, for simple data transfer, even mismatched impedances (e.g. 110 ohm to 75 ohm) usually are not a problem for short cable runs, as long as they do not cause glitches or dropouts. To reduce jitter issues, you should properly terminate all connections at the receiving end. This simply means using the right cable and connector (no y-cords!), since the termination resistor is built into the receiver. Using balanced digital connections reduces RF radiation into sensitive analog stages compared to coax cables, although an analog input should be immune to digital interference if it is well-made.

There is only one right "kind" of digital cable and connector—one whose impedance is a correct match for the circuit (e.g., 75 or 110 ohms). In fact the traditional *phono* plug, also known as an *RCA plug*, commonly used for S/PDIF connections, is not a true 75 ohms, and the standard XLR connector, used for AES/EBU connections, is not a true 110 ohms. Some manufacturers have created special versions of these connectors that have the correct impedance. But generally the connector design is the least of our worries. Keep your connections short and avoid extensions, because connections in the middle can add some reflections to the circuit. You may not notice a problem until running a very high sample rate. Ordinary microphone cable is not recommended for AES/EBU; use a cable rated at 110 ohms. Some audiophile manufacturers have marketed cables that are completely improper for a digital circuit, but since they affect the sound of a typical consumer-grade DAC in unpredictable ways (usually by adding jitter, not reducing it), consumers have been known to play with such cables to "tune" their systems. This is a losing battle, because cable-induced jitter reduces resolution and colors the sound.

The Internet and Jitter

The Internet has no clock. "Realtime" files played over the internet pass in irregular packets; they meet a clock for the first time when the computer gets ready to feed a DAC. The key to clean Internet monitoring is to use a large enough buffer, followed by a crystal clock and a jitter-immune DAC.

V. Mixing, Processing And Jitter

Jitter does not affect the data in...

• a real time all-digital mix in digital consoles or DAWs. After the initial A/D conversion, the data can pass from processor to processor, from medium to medium, regardless of clock jitter — just as long as the interface jitter is low enough to allow an error-free transfer. Similarly, clock jitter has no effect on the performance of outboard digital processors, which are all *state machines*. A **state machine** is defined as any type of processor that produces identical output for the same input data. It does not look at data timing or speed, but only at the state or recent history of the data. Digital processors look at the value of the samples, not their time of arrival. In other words, digital processors are completely immune to jitter. We could make the clock completely irregular, or even slow it down to 1 sample per second (assuming the processor could lock to such a slow signal), and eventually the processor would output all the correct data words. The only exception to these rules would be a processor or DAW that contains an ASRC, which *does* look at incoming and outgoing clock while it resamples the data. An ASRC is not a state machine because it produces different output data on successive passes.

• a non-real-time situation, e.g., when any processor takes in a file and outputs another file, operating without a clock, jitter has no meaning. When transferring from file to file (so-called *bounce to disk*, digital bounces or captures), the process deals with one sample after another. The data file gets stored sequentially regardless of clock. Consider this "bowling ball" analogy: throw a series of bowling balls down the alley, some white and some black. Although their timing is irregular, when they land back on the stand, the white and black balls are in the same order, so the output data is identical. If audible differences exist between capture and source, it's likely an issue with automation not keeping up, or a slow CPU not managing all the calculations in the required time.

"Most digital processors are completely immune to jitter."

• communication between plugins, which is asynchronous, happening in fractions of real time. Data is stored in a buffer at the end of the plugin chain for real time output when it is clocked out of the buffer.

All current professional oversampling processors are also state machines, because they use synchronous SRC, which itself is a state machine.[8]

Although some digital pitch processors such as Autotune™ are not state machines (due to their randomizing algorithms), these too are not affected by jitter. They deal with each sample coming in, one at a time, regardless of the regularity of the clock feeding the box.

Jitter affects the monitoring

Jitter becomes meaningful in a digital mix only during monitoring, when the data is clocked out of a DAC. This is where everyone gets confused. Let us emphasize: if high jitter during the monitoring *does seem* to affect the overall sound quality, it really only affects that individual listening experience, and has no effect on the data. This is what I call "ephemeral jitter." As Andy Moorer says, *don't confuse the messenger with the message! The message (the data) remains intact; so if it sounds degraded, blame the messenger (the clock inside the monitor DAC)*. If connections are improved and the sound gets better, it does not mean that the digital equalizers are suddenly performing better — just that a cleaner clock is getting to the DAC.

Jitter affects the data during a digital mix only...

- when signal leaves the digital realm to use outboard analog processors. Hence superior converters and clocking must be used for outboard equipment feeds.
- when using a digital console containing ASRCs, which are not state machines. Many digital consoles contain cheap ASRCs, which have low resolution. All ASRCs affect the data, are sensitive to clock jitter, and are especially problematic in low cost consoles with compromised clocks.

Analog Mixing

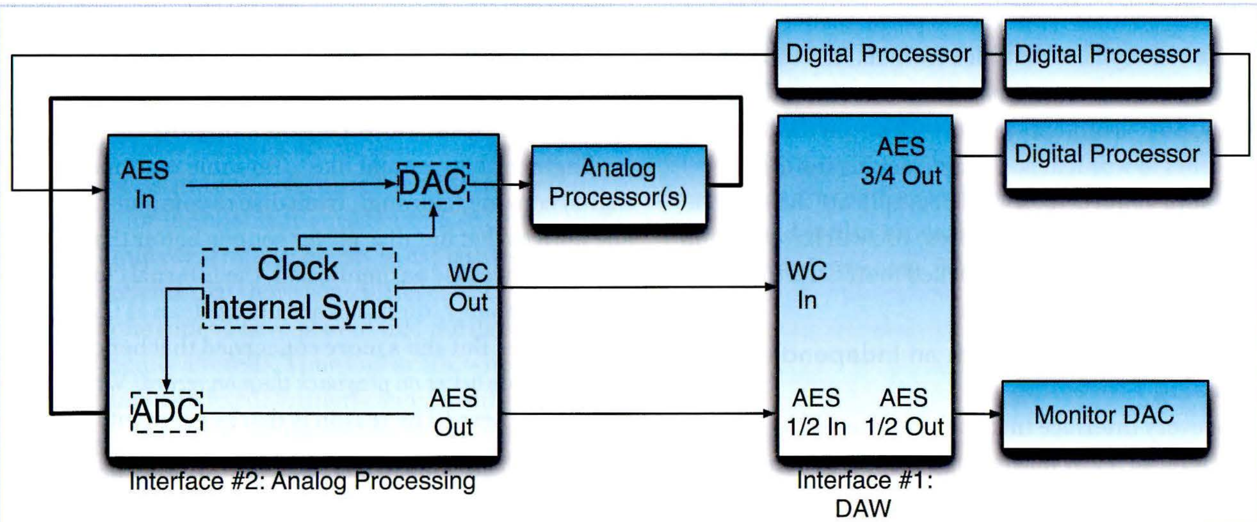
Jitter performance is critical when mixing from a digital source (e.g., DAW) with an analog console. If you have invested considerable money in an analog mixing console, it pays to investigate whether internal or external clock is best for you, and what kind of external clock. When in doubt, use internal sync: it's the safest option.[9]

VI. Real World Examples

"Don't confuse the messenger with the message."
— Andy Moorer

Example A: Mastering with External Processors

Pictured below is a block diagram of signal and clocking in a mastering studio. The DAW is attached to Interface #1, an interface with digital inputs and outputs, so we use Interface #2 for analog processing. We want the sound of the analog processing to be its best so we choose a superior interface that has an ADC and DAC that are locked internally to its low-jitter crystal clock and whose PLL performance does not improve with external clock. Interfaces like these incorporate a low-jitter, high-frequency clock connected via short PC board traces to the clock input of the converter chips to produce the lowest jitter and highest sound quality. This clock becomes the master clock for the entire system. The weak links are the phase-locked loops — for example, the word clock PLL in the word-clock input of Interface #1. The other weak link is the



Mastering system block diagram. Interface #2 contains the system master clock. It internally stabilizes the processing ADC and DAC, sends and receives signal from the analog processors. Any jitter in the digital processing chain is irrelevant, since the ADC and DAC use the stable internal clock of Interface #2. In other words, the jitter of the digital processing chain has absolutely no effect on the converters.

PLL in the monitor DAC, which must reject interface jitter coming from AES 1/2 out. This means the clock jitter in Interface #1 and the jitter rejection of the DAC are important to the monitoring quality, but to nothing else! You might ask why we do not use another DAC in Interface #2 for monitoring, and we could — but the system would not be as versatile. The monitor DAC we use is an independent DAC that can slave to "foreign" rates and sources, e.g., CDs, Blu-Rays, or other DAWs; this gives us the versatility of listening to any source at any time. For the sake of versatility, we accept a possible loss of quality by using a DAC that slaves to its input, so it's important to use a monitor DAC with excellent jitter rejection. Grimm Audio has a solution for "jitter perfectionists" called the CC1 master clock, which can clean up a word clock and clean up and relock an AES/EBU feed, which could be used to improve the sound of the monitor DAC.

Since AES/EBU jitter is generally greater than WC jitter, you might ask: why not lock the digital processors to word clock? The reason is that jitter in the digital processing chain is irrelevant to performance, because Interface #2 replaces the clock from its AES input with its own master clock, which drives the DAC chip directly. So it is unnecessary to lock any of the digital processors to wordclock. In fact, locking the digital processors to wordclock would complicate the system: whenever changing sample rates, you will have to jump through hoops. Gear that is locked to AES/EBU will automatically change its rate.

Example B: Mastering with an Independent DAC for Analog Processing

Not every interface incorporates mastering-quality converters, so many mastering engineers choose to use a pair of standalone high-quality converters. The standalone DAC will slave its clock to the incoming AES/EBU signal (there are very few DACs with wordclock inputs). Despite the fact that AES/EBU connections normally are a bit jittery, design and mastering engineer Eelco Grimm has proved that a DAC-ADC chain includes a "free ride," by showing that jitter distortion mostly cancels out when the clock is the same source for both, which is the case here. This means that the ADC will largely cancel out the jitter-induced distortion from the DAC. Phase shifts in equalizers and other analog processors will reduce this "free ride" to some extent. Someday I will perform a shootout of these options, but in any case I find that converters have gotten so much better that with simple attention to the details in this Chapter you will get excellent sound no matter which option you choose. Eelco's discovery explains an experience I had during a location recording session: I heard better sound while recording than when playing back, because during recording, the DAC is locked to the "live" ADC, so any jitter in the ADC is cancelled out. But on later playback the DAC was slaved to another clock, and we can now hear the distortion engraved in the medium, caused by the jitter in the ADC.

Example C: Real Time Digital Copying

Engineer Betty would like to do some realtime digital copying (cloning), from disc to computer. First she notices that her disc player sounds better than her computer because as mentioned, the internal clocks of typical computer interfaces are not as clean as those in disc players. But she's more concerned that her computer sounds *better on playback than on record!* What is going on here? The reason is that her computer's internal oscillator is performing better than its PLL on external sync.

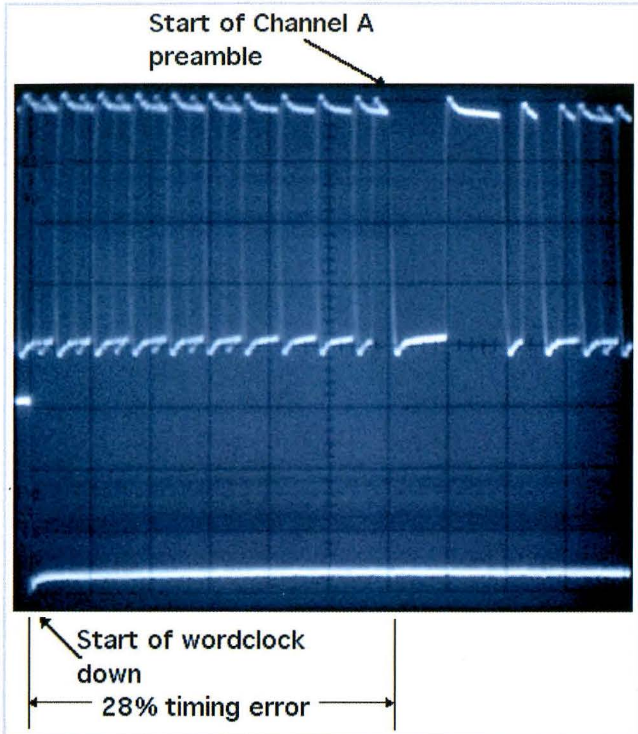
Digital copies really are perfect. Betty asks, "shouldn't I recopy the file now that it sounds better?" But as this is a case of ephemeral jitter, Betty's data is just fine and should sound its best when it is properly clocked.

Example D: I Hear a Difference Copying via S/PDIF vs. AES/EBU

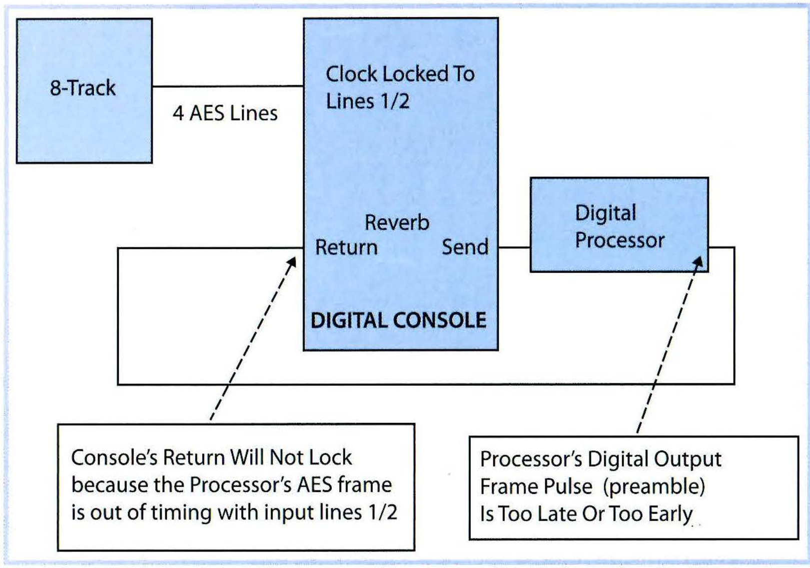
Engineer Don believes that AES/EBU sounds better than S/PDIF through his DAC. So he decides he should make all digital dubs through AES/EBU, which is simply the wrong conclusion. All interfaces are designed to pass data unaltered. However, the DAC which is monitoring these interfaces may be sensitive to incoming jitter and one interface may exhibit more jitter than the other, thus leading to Don's mistaken conclusions. Even if he doesn't replace his DAC, he can safely make digital copies through either interface. He can prove that both interfaces are equivalent by doing a null test on two consecutive digital copies (see Chapter 22).

VII. Concern for the rest of the world...

Since most consumers listen to music on DACs which are more susceptible to jitter than pro gear, it's very important that the references we make for our clients have the best possible sound. There is no jitter on a storage medium, but clients hear differences, and complain when their pressing doesn't sound as good as their CDR reference or their reference file. Ultimately the solution is for the clients to get a better DAC. Until that time, it is quite frustrating for me to explain to non-technical clients that there is no difference even though they hear one (once I have proved that the data has not been accidently altered). Time and again, when the clocking has been fixed, formerly audible differences disappear. Furthermore, we can magically "restore" the sound quality of an "inferior" CD by copying it back to a workstation and then asking the client to play this file next to their reference file (that's the best proof I can give to my client, though this whole procedure wastes so much time). In this case the digital dub can sound better than the original CD! But the two files played next to each other should sound and measure identical. No wonder Chico and Groucho are confused. Fortunately, these issues are tending to disappear with the advent of file-based media, since the client hears everything played through the same DAC with the same clocking.



This oscilloscope photo compares the timing of the start of the Channel A AES preamble against the start of wordclock at the output of a digital processor. This timing offset of 28% of the length of the AES frame is 3 points greater than the permissible tolerance in standard AES11 and would cause locking trouble to intolerant consoles or DAWs or other receivers.

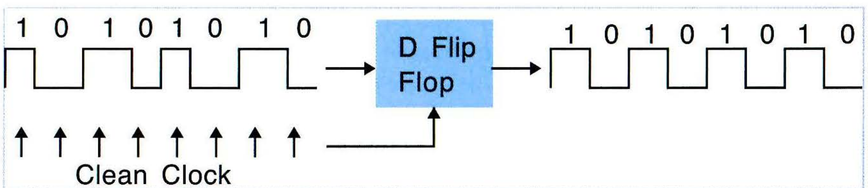


AES to AES framing error can cause locking problems

VIII. Things That Go Bump In The Night

Framing and Timing Errors: Wordclock to AES timing error

Although jitter is often the scapegoat for a motley of problems in a complex digital system, the fact is that 99% of the time, glitches, clicks, dropouts, noises and lockup problems, are caused by *framing problems*, not by jitter at all. Framing problems are caused by timing differences in critical signals and cannot be solved without software or hardware modifications. Pictured (page 333) is an oscilloscope photo, at the top of which is the start of the AES preamble (which defines the beginning of the AES data word), and on the bottom, the point where wordclock changes from high to low.



A Simple Reclocking Circuit

To complicate matters, there is no standard that defines the synchronization of the wordclock transition (low to high or high to low) with the AES preamble. The choice of using low or high is a timing difference of 180 degrees, or approximately 11 µs at 44.1 kHz, which is enough to cause glitches, or lose signal completely. One model of workstation has a menu choice that allows us to choose the wordclock phase, making it more compatible with products of various manufacturers. But the best solution is to use AES/EBU as a clock source, ensuring that all clocks will be in phase. However…

AES to AES framing error

Digital audio is a small industry, still experiencing growing pains. And since some digital audio processors produce an AES output that is out of timing with their AES input, intolerant consoles and workstations have trouble locking to them. Once I was forced to insert a simple reverb unit via analog, because the digital console would not lock to it on a digital send/return path. The fault was caused by the console's intolerance to AES framing errors, aggravated by the reverb unit's output being slightly out of framing (timing), as seen in this figure (pictured above left). We can probably prove it's a framing problem without an oscilloscope: in this situation, set the external processor to run on its internal clock, and lock the console to the external processor on its reverb return. If the console will lock and pass audio from the external processor, then the previous problem was due to framing issues.

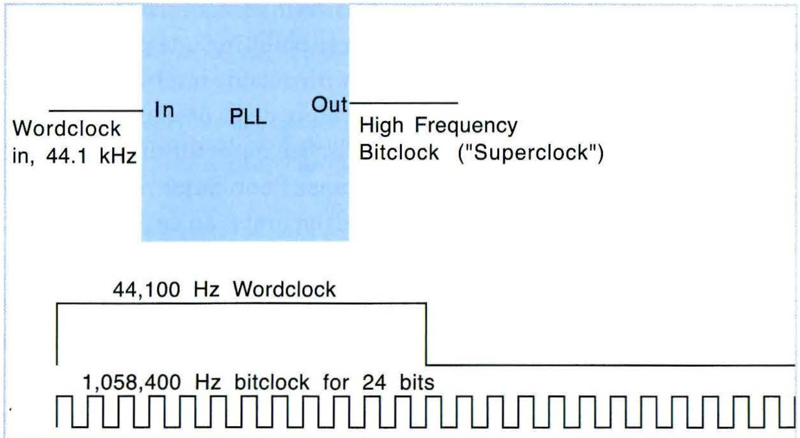
Framing errors are cumulative in a chain of processors if they are chained via AES/EBU (or SPDIF). If the framing error of each box is in the same direction, then the total error could be enough to cause locking problems in sensitive consoles and DAWs. We may be able to stabilize the system by locking the last processor in line to external sync (wordclock or AES). If the last processor in line is framing-tolerant on its AES input, then locking it to external sync will force its output to a known framing and hopefully to within the tolerance of the DAW.[10, 11]

IX. How It Works

Simple in Theory...

Most engineers don't need to know the technical details of how equipment works, but usually a couple of nagging questions remain, like... What is a PLL? Is it the same as a reclocking circuit? Why do we need a high-frequency clock?

Reclocking Circuit. The data inside typical audio processors is moved along by a clock pulse, traveling serially from chip to chip. This clock bus is distributed to all the critical chips inside the box. As we've seen, even if this clock is jittery, proper data still makes it to the next chip in line. But sometimes the clock signal needs to be *reclocked*, for instance to have the lowest jitter when feeding a DAC. Pictured (page 334 bottom) is a simple reclocking circuit; on the left side is an incoming data word that's been clocked by a jittery clock; the data value is (conveniently) 10101010. This word passes, one bit at a time, into a logic circuit called a *D-type flip flop*, which is being fed a clean clock. Almost magically, the data neatly marches out of the flip flop, and in theory, all the jitter is gone and the data is ready to feed the DAC in perfect time with the clock signal. Notice how the clean clock's pulses permit the flip flop to properly "sample" each data value, but only if the clock pulse lands within the acceptance time of each incoming bit. In this illustration, the fourth (and eighth) data bit is in danger of being missed if it arrives a moment later, in which case the clean clock would land on the previous bit and the wrong data would be output, producing audible clicks or glitches.[12]



A PLL is needed to generate the higher frequency clock required to move the individual bits from place to place.

The figure (above) illustrates why a PLL is needed. If we are passing 24-bit audio bit by bit, then we need a high-frequency clock pulse that is 24 times the frequency of wordclock. Wordclock enters the device, and has to be multiplied up to the higher frequency to drive those bits around, known as the bitclock. It is easy to divide down without creating jitter, but very difficult to multiply up, and it's the job of the sophisticated circuitry of the PLL to create the higher frequency while reducing incoming jitter.[13] A PLL-based circuit is a sort of electrical flywheel: it tries to find a center, holding reasonably steady while still following the average frequency of the incoming source.

...Complicated In Practice

What makes these circuits so difficult to design is that in addition to developing a perfect PLL, at high frequencies, leakage from some portion of the circuit can travel through back paths to contaminate the clean portion of the circuit. These paths include power supply and ground. Couple that with outside interference and ground loops, and you have created a designer's nightmare. 10 picoseconds error can make the difference between an 18 or 20-bit noise floor. Some manufacturers use a dual-PLL, where the first is an analog circuit, and the second a voltage-controlled crystal oscillator (VCXO), in an attempt to get the jitter down to that of a quartz crystal. Unfortunately, designs using VCXOs cannot varispeed because of their narrow frequency tolerance. It is difficult, yet possible to design a jitter-immune PLL that's as good as a crystal, has wide frequency tolerance and quick lockup.

In previous editions, I presented a number of jitter measurements using the J-Test signal invented by the late Julian Dunn, an independent consultant best-known for his work on the Prism brand of converters. But I am now confident that the best clock equipment designs have exceeded the resolution of the J-Test regarding the important characteristics of jitter using the analog output of converters, except for broken gear. Likewise, I once built a clock jitter analyzer whose 100 ps performance was soon exceeded by the quality of the best gear. So at this point I depend more on my ears since I cannot afford an analyzer with sufficient resolution! Quite simply: If the sound is wider, warmer and deeper, chances are the jitter is lower! If you would like to make simple, low-resolution jitter measurements, two possibilities remain for us poor mortals:

1) Use an FFT analyzer to compare high frequency harmonic distortion with low frequency. If the 20 kHz HD reads significantly more than the 1 kHz HD, then jitter is the likely cause. Try to compare like with like, for example, a reference converter with an unknown converter.

2) The J-Test, which is only relevant to AES/EBU or SPDIF interfaces. Compare the level of the spikes between two different converters under test. If you find a converter with lower spike levels for all the different jitter frequencies, this is probably the superior converter.

In Conclusion

Your clocks have been set, they are steady. Now it's time to run them and enjoy better sound!

1
I still prefer the version I misremembered in the second edition of this book. It's much funnier: Margaret Dumont catches Groucho embracing a beautiful woman. In defense, Groucho quips, "Are you going to believe me, or your own eyes?"

2
Leading to the "Wordclock Du Jour", every one sounding different and maybe one of them is right, or none! Related: an ignorant audiophile magazine DAC review marveling at a DAC that's "good enough to reveal digital cable differences!" Fact is, a device that "reveals" an apparent difference must be considered *defective*.

3
Dunn, Chris & Hawksford, Malcolm (1992). Is The AES/EBU/SPDIF digital audio interface flawed? *Journal of the AES* preprint 3360.

4
CD medium testing includes a rather esoteric and confusing measurement called *jitter* which has nothing to do with clocking.

5
According to a simplified formula: Moses, Don (October 1992) Enclosure Detuning for 20-Bit Performance, *Journal of the AES preprint 3440*.

The following expression utilizes Carlson's similar triangle analysis method and is useful for the case where: (1) the jitter deviation is small compared to the sampling interval, (2) distortion is measured at the zero-crossing of a sine wave, (3) the peak-to-peak amplitude is normalized to 1-V, and (4) the maximum slope is approximated as 2 x the information bandwidth:

Resolution (in dB) = 20 log (time deviation x 2 x information bandwidth)

For example, 25 ps of jitter, 20 kHz information bandwidth, yields:

20 log (25 ps x 2 x 20 kHz) = -120 dB, which provides 20-bit resolution.

In other words, if we double the sample rate to 88.2 kHz (the information bandwidth becomes ~40 kHz), the same amount of jitter reduces signal to noise ratio by 6 dB. For 20-bit performance, at 88.2 kHz, if we consider the information bandwidth goes to 40 kHz, the jitter would have to be halved, to less than 12 picoseconds. And for each 6 dB improvement or 1-bit increase in wordlength, the jitter must be halved again. Even if we limit the information bandwidth to 20 kHz, in order to get excellent performance with long wordlength, it boggles the mind the degree of engineering care required to produce a low-jitter converter.

6
In a multiconverter situation, which box should be the master? According to Ian Dennis, in Resolution Magazine Oct 2006:

Using the best box in the studio is usually misguided, since good boxes are equally at home as master or slave whereas bad boxes are usually OK as master but perform poorly when slaved.

7
Crystal Semiconductor application note, A/D Conversion with Asynchronous Decimation Filter.

8
Any processor which adds random dither is not a state machine in the strictest sense. All good synchronous SRCs employ internal dither to linearize the process if outputting to fixed point. The randomizing effect of dither means that each output pass will produce slightly different data *at each instant*. However, *on the average*, the output stream is really the same, and if we could subtract the random dither from the output signal, each pass would be identical.

9
In the vast majority of converters manufactured today, "the jitter caused by wordclock is typically 15 times higher than when using a quartz based clock", according to the manual for the RME model ADI-8-DD format converter.

10
The best hardware solution I've found to framing issues is an external rackmount AES/EBU interface made by RME. Its framing tolerance is excellent; it cleans up framing issues, conforms all inputs to the same framing on its output, saving the day by making unrelated sources look like a single digital multitrack to my sensitive DAW.

Julian Dunn clarifies:

Framing is a synchronization issue covered in AES11. These define the permitted output alignment error (+/-5% of a frame period) and the tolerance to input timing offset (+/-25% of a frame period) before the delay becomes uncertain. The specifications for the interface itself (AES3, IEC60958) do not allow a receiver's ability to decode data to depend on the relative alignment of clocks—as long as the dynamic variation is within the jitter tolerance spec. (about +/-4% of a frame period at low jitter frequencies).

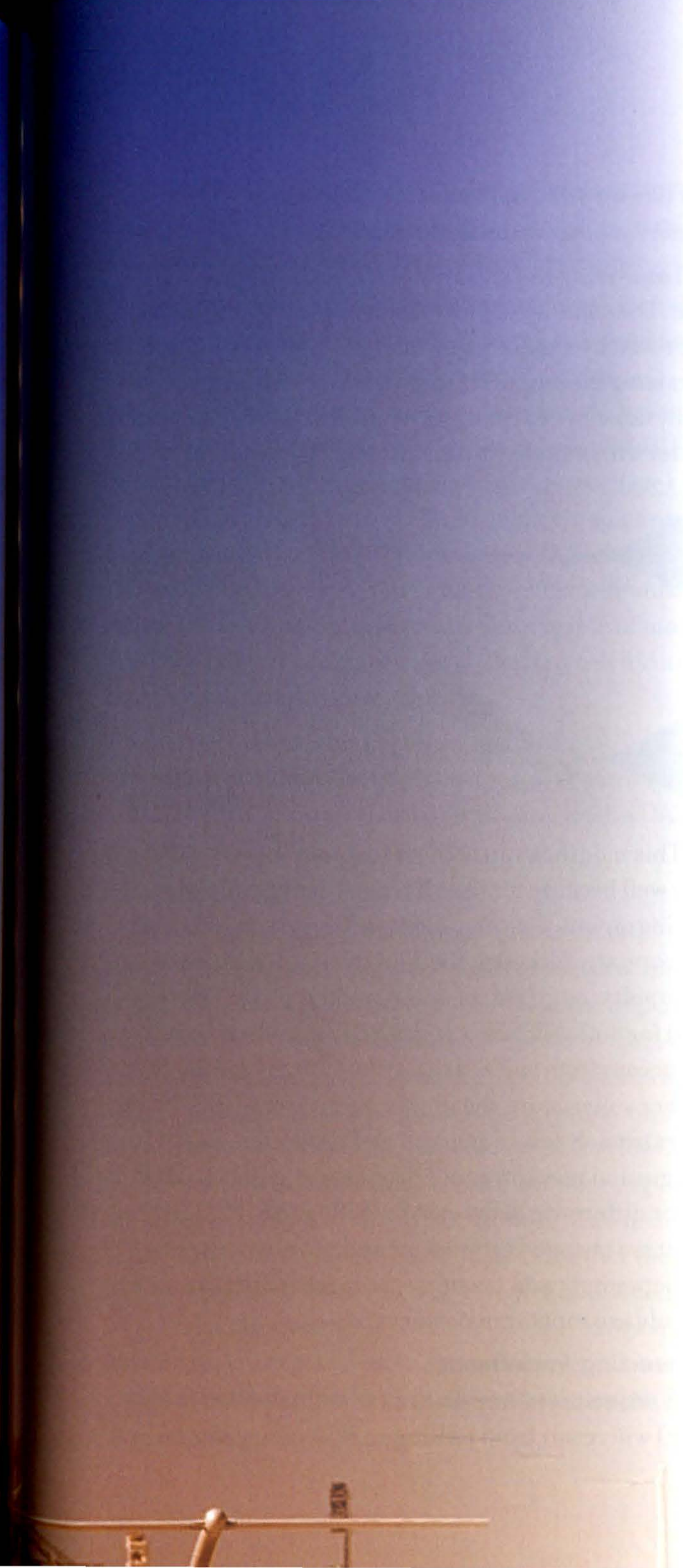
11
If you're spending $30,000 and upward on a digital console, request the manufacturer to sign an agreement that the digital inputs and wordclock framing tolerances must meet or exceed the AES11 synchronization specs or the manufacturer will correct the problem at no charge. This amounts to a sad wakeup call to the manufacturers, but consumers should be entitled to interface real-world equipment to their consoles.

12
Those dyslexics in the audience will appreciate that I am taking slight liberty with this discussion for ease of understanding. Since the left hand end of the bitstream is the last to go into the flip flop, the "fourth bit" counted from left to right is actually the fifth bit to go in! This leads to the requirement that software has to decide whether to make the left or right end of the bitstream be the most significant bit. Intel and Motorola have been fighting over which end comes first for decades.

13
Many bitclocks are 32x the wordclock, or greater, to allow for a longer internal wordlength. A typical PLL may generate a superclock which is 128, 256 or even 384 times the wordclock frequency, and is then divided down using a simple divider.

14
However, a crystal which is significantly off the standard center frequency can cause locking problems, if using a low-jitter PLL with a narrow lock frequency range (which are also intolerant to varispeeding). But if the system components lock, then an off-standard crystal won't affect digital dubs at all. Some PLLs have a narrow and wide setting to deal with sources that are a bit off the standard. Switching to wide increases frequency tolerance, but also increases the PLL's jitter. Don't be concerned, as long as the PLL is not driving a converter. If a system has locking problems not due to framing errors, measure the source sample rate, and if it's off tolerance, trim the master crystal oscillator frequency.


Discovery



CHAPTER 25

Technical Tips and Tricks

Introduction

I'd like to conclude this edition of *Mastering Audio* with some tips on how to maintain a digital audio studio.

I: Debugging Digital Interfaces

When the AES/EBU and S/PDIF[1] interfaces were created, the idea of using standard microphone connectors and cables seemed like a godsend, but these were never intended to carry the high frequencies of digital audio (about 6 MHz bitrate for 48 kHz SR). So, eventually we ended up with special RF-rated cables attached to our old-fashioned XLR connectors. However, AES/EBU cable makes excellent analog line and mike cable due to its high bandwidth and low capacitance.

ADAT not recommended: Technical guru B.J. Buchalter has pointed out that the ADAT interface has a tremendous amount of jitter — 80 ns on certain devices. The result is that data transmission errors often occur, and the commercial receivers go to great lengths to engage in error concealment. A second problem is that, since ADAT does not carry a sample rate flag, we have to jump through hoops whenever we switch between single and double sample rates. So when possible, for real time interfaces, choose MADI, AES/EBU, even S/PDIF or Toslink in preference to ADAT.

Software Issues

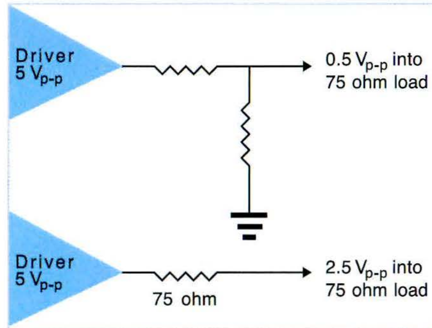
Sample Rate Flags: When recording from an external digital source, most DAWs ignore or cannot read the incoming sample rate flag (metadata). It is up to the user to manually set the session to the correct sample rate. For example, a session might be set to 48 k when it is actually receiving 44.1 k. Occasionally, the DAW will assign the wrong sample rate flag to the file it creates, and the file later plays at the wrong pitch. Fortunately, the audio data is still correct, so the solution is not to panic: fix the flag in the file header (metadata), which can be performed by *Soundhack* or *Sample Manager* (both Macintosh programs) without having to rewrite the entire file, very useful for batch processing errors of this type.

Hardware Issues

Glitches or intermittent sound are most likely hardware or interconnect issues. We don't know how well a digital audio receiver is working until it stops making sound or produces audible glitches. This is why digital audio receivers ought to include signal-quality indicators. The proper way to assess a hardware interface is to measure its objective performance by looking at the width of an eye pattern on an oscilloscope. Always use matched-impedance cabling, especially for long runs and high sample rates. When measured with a terminated load, the 110 ohm balanced signal should measure between 2 and 7 volts p-p, while the unbalanced signal should not be below 0.5 Vp-p.

Shields are unnecessary in the *balanced interface*, as can be illustrated by the success of Belden's Mediatwist™, consisting of four bonded-twisted pairs for up to 8 channels, which performs as well as the highest-grade coax. In fact, standard Cat 5 or 6 twisted pair Ethernet cable can carry four AES/EBU signals (eight channels) very well. The biggest problem with the unbalanced consumer interface is that its low voltage (0.5 Vp-p) does not give much margin for error with the losses of coax cable. These issues could have been avoided if the S/PDIF interface protocol had specified 1 volt like the **AES-3ID** standard, which uses a BNC connector, popular with video houses.

Improving the stability of the unbalanced interface. The stability of the unbalanced interface can be improved by upgrading to *low-loss* 75 ohm cable, and/or by raising the output voltage from 0.5 volts to 2.5 volts, easily done by replacing the voltage divider at the transmitter with a single 75 ohm resistor, pictured here:



Voltage level improvement for S/PDIF transmitter

This modification to the transmission side works very well because it raises the noise margin of the receiving circuit. *Warning: modifying circuits usually voids the warranty.* Although the AES standard is between 2 and 7 volts, note that the same audio receiver chip is used for both AES/EBU and S/PDIF decoding, and it can accept from as low as 200 $mv_{p\text{-}p}$ to as high as 7 volts. Higher voltages are usually not a problem with S/PDIF, but extremely low source voltages reduce the noise margin and may introduce dropouts or glitches. The major difference between AES/EBU and S/PDIF at the input is a change of connector and termination resistor between 75 and 110 ohms, as most S/PDIF circuits already use input transformers.

Converting Impedances

A mismatched impedance (as well as circuit imbalance) will result from putting an RCA connector on one end of an XLR cable without changing source or load resistors. However, short cable lengths and low sample rates may adequately pass the mismatched signal.

An impedance-matching transformer is definitely the best way to convert between 110 ohm balanced and 75 ohm unbalanced. In past editions I had recommended a cheap resistor-based solution, but I have since found that, especially when going from S/PDIF to AES, many senders are substandard (low voltage) or receivers are insensitive. So the resistive solution fails, especially on long cable runs or with high sample rates. Therefore, if you want a solution that will work no matter what situation, and you are not capable of modifying the S/PDIF transmitter as above, buy a dedicated impedance-matching transformer.

Be aware that reversing pins 2 and 3 of an AES/EBU cable does not affect the audio in any way. Polarity reversal of the audio signal can only be done with a DAW, processor or console.

Cable Lengths

With copper cable, the higher the sample rate, the shorter the tolerable cable length, because of the possibility of interfering reflections from the impedances and connectors at each end of the cable. The AES3 standard specifies usable lengths up to 100 meters at 48 kHz, which is possible with careful termination and high-bandwidth, matched-impedance cable. However, at 1/4 wavelength, reflections are at their worst, aggravating errors with cables that are close to 20 meters (66 feet) at 48 kHz, or 33 feet at 96 kHz. Neither the XLR nor the RCA connector was designed with exacting impedance specifications, so avoid passive hardware patchbays, splices, and extensions. Cable length has an insignificant effect on latency, since a cable length difference of over 200 meters would be required to exceed the AES-11 framing tolerance.

With optical cable, the main concerns are bit integrity and interface jitter. As we explained in Chapter 24, jitter is only a consideration when performing a conversion. Plastic Toslink cable is rated for up to 5 meters (10 meters with some receivers) beyond which there is unacceptable signal loss. A legitimate test for cleanliness of an optical interface is *margin distance before dropout*. While observing the lock indicator on an AES receiver, or simply listening to the audio, disconnect the cable from the input and slowly pull it outwards. The amount of distance before losing lock is an indicator of the margin of sensitivity and the strength of the optical signal. 1/8 inch to 1/4 inch (6 mm) is a good margin. Glass fiber has much less loss than plastic, and can transmit for thousands of feet.

Two-wire 96 kHz and 192 kHz

Most current copper AES/EBU interfaces can carry two channels at or beyond 192 kHz sampling. However, if you have a DAC or other device that requires a "two-wire" or "S/Mux" connection, two cables will be required. In that case, one cable carries the left channel and the other the right. In such cases, listen carefully to ensure you have received the correct stereo signal. Keep good documentation and carefully mark cables to avoid reversed channels. Optical (ADAT format) can also run S/Mux: when the interface says 48 kHz/8 channels it's actually 96 kHz/4 channels.

II. Timecode and Wordclock in a Digital System

Drifting drifting drifting

A common task in a digital audio studio is to slave a sequencer via timecode to a master DAW. If the

> "There must be only one clock master in any system."

sequencer is *not* recording or monitoring digital audio, it can be set to timecode sync. It will then slave (adjust its speed) to the incoming timecode, and *will always stay locked to timecode. It will not drift.*

However, a slave machine or DAW playing digital audio must be locked to the same digital audio source as the incoming timecode generator, or it will *drift out of sync*. This is because the slave machine or interface takes a *timestamp* or *trigger* from the first valid timecode it sees. From that point on, it ignores incoming timecode and creates its own timecode locked to the digital audio clock. Also, make sure the slave is set to the same sample rate and timecode as the source.

A multiple sample rate system can be locked together in real time without causing timecode to drift as long as a real-time *synchronous* SRC is used. For example, when coming from a 48 kHz source DAW to a 44.1 kHz destination (slave) DAW. Even though the slave is set to a different sample rate, the frames are synchronous. AS-RCs are not suitable for this purpose, the second sample rate clock must be synchronously derived from the first.

Pull-ups and Pull-downs

If a computer is playing video instead of a standalone video deck, life is easy, even in NTSC countries. But when video decks are in use, pull-down sample rates may be needed. The digital audio output of a professional digital video deck can be substituted for a wordclock generator, because it will have the correct sample rate related to the video sync rate. A good resource for video and sync issues is Tom Holman's book, *Sound for Film and Television*.

Wordclock Voltages

The problem with standards is there are so many of them! With Johnny-come-lately digital, no voltage standard was developed for wordclock, which produced a chaotic situation. Many of the earliest wordclock generators were based on video sync (blackburst) generators, which produce 4 volts peak-to-peak into a 75 ohm load. Later, wordclock generators appeared based on the video standard of 1 volt. Yet a third standard is based on TTL-level, with 2.5 volts terminated, and 4-5 volts unterminated. Chances are that any wordclock lock issue can be traced to these voltage incompatibilities. The best solution is to use generators that produce 4 volts and receivers that can accept anything between 1 to 5 volts, which are cross-compatible. We may be able to fix a disfunctional receiver by removing the load termination resistor. If this doesn't work, then we need to measure amplitudes with an oscilloscope and get creative with the circuitry inside our distribution amplifiers. This may require removing the buildout resistor on one or more outputs in order to get a little more voltage. Caveat emptor.

No daisy chain: I do not recommend daisy-chaining wordclock, unless you are certain that the device with WC input and WC output does not add a delay on its wordclock output. Ordinarily I use a wordclock distribution amplifier, or a BNC-T connector at each wordclock input, with the terminator turned off except at the end of the line. See links.

III. Miscellaneous Computer Tips

Media Players

iTunes (OSX) and clocking: Repeat after me — iTunes is a consumer-based application. iTunes is a consumer-based application. Apple has made iTunes foolproof, it's a wonderful, simple-to-operate application in the typical Apple way. But audio professionals need to manipulate things that iTunes normally keeps "under the hood." We require:

- a media player that does not sample-rate convert in the background whenever the interface and the audio file have different rates
- a media player that controls and changes the clock inside the interface when switching between files of different sample rates
- a media player with user-selectable dither, including 24-bit dither to feed the DAC at its proper resolution
- a media player that can play all file formats including FLAC [Flac is a lossless audio file format]

To satisfy all of the above requirements, I nominate JRiver, Pure Music, Amarra, or Audirvana as alternative media players to iTunes. Pure Music and JRiver are my media players of choice. JRiver is available for Mac or PC. I believe the Mac version currently does not supply 24-bit dither, but the PC version does.

Best timesaver: One OS addition available for Mac (an equivalent is available for PC) is so essential that I'm puzzled OS designers did not build them in. I am referring to an addition called *Default Folder X* (Mac) or *Direct Folders Pro* (PC). These additions allow us to easily navigate to recently-used folders and favorite folders. Since I am switching from project to project throughout the day, I wager they save at least 1/2 an hour per day and countless mental frustration.

Another great timesaver: I nominate *Better Finder Rename* (OSX), a batch file renamer, for dealing with audio file names. This also works for files created on the PC side, since our audio server mounts on all client computers. For example, I may produce a bunch of intermediate files whose names are "I Love You Dearly 3244 mast.wav", "Looking Up 3244 mast.wav", etc. I bounce them with 16-bit dither to new files, but their names are unchanged. Better Finder Rename lets me quickly replace all occurrences of "3244 mast" with "1644 mast", which is a big help.

Another tip: Some audio programs may not produce audio if the file name includes characters with accent marks or foreign characters (which is not good design, but worth knowing).

Printing rant: A big thank you to whoever invented printing presets on the Mac and a big *what the hell* to whoever did not implement these timesavers on the PC. I am constantly changing my printing needs, e.g., color to black and white, dual-sided (duplex) or single, printing on a form from the second tray or from plain paper in the main tray and—you name it. The PC is just a time-waster, costing us anywhere from 2 to 5 minutes per printout to make sure the printer is properly configured, and often forcing us to dig through many driver screens. On the Mac, printing is not a process, it is a dream, taking only seconds using the printing presets pull-down menu. Adobe: *what were you thinking?* Acrobat bypassed OSX's excellent printing system, requiring my constant attention every time I want to print a PDF a certain way on the Mac, and wasting at least a minute (or bad printouts) each time I print from Acrobat on the Mac, because you did not implement printing presets in your application.

IV. Lightning Protection

Telephones: Florida is the lightning capital of the world. Lightning blew three successive expensive telephone CPUs and a number of extensions (in three separate nearby lightning strikes at our studio), because telephone wiring acts as an antenna for the electrical impulse from lightning. I did find some telephone surge protectors, but in the end I converted to a wireless telephone system. Problem solved.

Ethernet: Ethernet is extremely vulnerable to lightning, which has blown out Ethernet on computer motherboards and switches. Please don't write me about "proper grounding;" keep in mind the old riddle, "when is a ground not a ground?" The answer is "whenever lightning is involved." A one-meter power cord is enough to produce dangerous differential between ground and the Ethernet router chassis when lightning is involved. The only solution I have found is initially expensive, but it pays for itself in the long run: Ethernet protectors from *Furse*. I use the ESP Cat 5 Range for long cable runs and (to save on cost), the ESP LN Series for short cable runs. Since installing these, I have not had a single Ethernet device go bad. The key is to ground the protector's strap to a chassis screw on the device you are protecting, and keep it short: do not extend the ground strap. That way the protector's "ground" to which it dumps lightning current is at the same potential as the chassis of the Ethernet switch. You don't need "true ground," what you need is to "ground" the protector to the device you are protecting. This works every time! Once time someone forgot to attach the strap, and the next time lightning struck, the protector catastrophically went up in smoke, as did the Ethernet switch. Some cheap Ethernet switches do not have chassis screws — I have soldered the strap to the metal surrounding an Ethernet jack. I hope this information saves some of you a lot of trouble!

Power: There is only one good solution to power surges: a renewable, series-mode protector. It does not self-destruct when it receives a surge, as would an MOV. The company I recommend is *Surge-X*. It is expensive, but it helps the audio sound better by not dumping leakage current down the powerline ground, and properly protects from surges. Enough said.

V. Analog Audio Tips

Logging Analog Compressor settings: One of the things that distinguish so-called "mastering grade" processors from "standard" processors is their repeatability. Many of the dials are expensive detented controls. But a number of desirable processors do not have detented controls and for revisions we need to return a compressor's threshold and makeup gains to the settings we used on the project. So I document the threshold and gain settings of non-detented compressors by using a sine wave test tone located in my DAW. I log two meter readings, one with the threshold off to set the output gain, and the other with the threshold adjusted to the setting I used for the session.

1
S/PDIF stands for *Sony Philips Digital Interface*, which grew into the IEC60958 standard, which supersedes IEC 958. Officially, type 1 is consumer with the consumer bitstream (protocol) on unbalanced RCA or Toslink optical connectors. Type 2 is professional, with the professional bitstream over XLR balanced connectors. There is also the AES-3ID standard, which transmits the professional bitstream over a 75-ohm BNC connector at 1 volt p-p.

Afterword

How This Book Was Written and Edited

After the production of the second edition of *Mastering Audio* the music and audio worlds underwent radical changes. When it came time to produce the third edition, I had created probably a thousand notes on how I wanted to present the new areas. I then integrated these notes, emails, maillists, website forum feedback, and seminar feedback into a holistic entity. I rewrote each Chapter with this outlook and incorporating the latest developments in audio theory and practice. The result is a fresh book which is refined, tightly woven, and completely reorganized to reflect an audio engineer's workflow and thought process. It's eminently suitable as a course curriculum for audio schools or a progressive story that a musician, producer, or engineer can read from start to finish.

During the rewrite, my new editor Chris Morgan and I exchanged semantic flourishes and fixed more than a few syntactic accidents. I am grateful for the contributions of previous editor Eric James, many of which remain intact in the third edition. No Author Is An Island!

Mary Kent is the designer for the third edition. The fresh new "cinemascope" book design is in glorious full color on glossy stock. Many new illustrations and graphs have been created and many old images have been re-created. Mary's photos evocatively complement and illustrate the audio concepts of each chapter.

Again, thanks to the many readers, correspondents, technical consultants and seminar audiences who have helped mold this third edition and inspired new ideas. I am thrilled with the way it has come out—this book flows in a warm, organic way. I hope the audio world goes through a lot less drama in the next 10 years than it underwent in the last 40!

Closing Thoughts

If you have or are planning to have a long career in audio, be sure to actively *preserve your hearing*: Wear earplugs (15 dB or more) while driving on a long trip to reduce ear fatigue, while flying (also helps reduce pressure changes with altitude), at loud concerts, and on noisy city streets. I hope you will see me around, and continue to hear me around!

Sound good!

Bob Katz

June 2014, Orlando, Florida

"No Author Is An Island!"
— Bob Katz

"we'll FIX it in the shrinkwrap."

—Frank Zappa

Appendix

Introduction

Some of the items previously in the appendix have been moved to online resources. As always, visit *www.digido.com/media/links.html* for all links mentioned in this book. There you will find a link to a powerful online conversion calculator hosted by Tonmeister Eberhard Sengpiel. Also, a bibliography with recommended reading, CDs for equipment testing, ear training courses, and much more.

Appendix I. The Art of the Album Sequence

Introduction

Sequencing an album is an art. It is possible to turn a good album into a great album by choosing the right song order. The converse is also true. Sometimes, the musicians making an album have a good idea of the song order they'd like to use, but many of my clients ask me for advice on ordering their album.

Traditionally, the label's A&R person would help put the album in order, but with independent productions that service is not always available and so it falls to the producer, or someone experienced, and politically "neutral." A neutral producer helps prevent the "more me" syndrome, and achieve the goal of a cohesive album rather than a collection of different band member's tunes. An experienced mastering engineer is well placed to provide useful guidance during this process. I am a musician, though not actively performing; my song-ordering expertise began in 1972 with my first free-form radio show on WWUH-FM, when I would construct the sequencing of an entire 3-hour progressive rock radio show. I learned to construct "sets," groups of songs that would go together based on a musical or intellectual theme.

My advice to album-builders is to avoid intellectualizing. One client wanted to order his album by the themes of the lyrics; he started with all the songs about love, followed by those about hate, and finally the songs about reconciliation. It turned out to be a musical disaster: The beginning of his album sounded musically repetitive, because all his love songs tended to use the same style. His progression of intellectual ideas was not immediately obvious to the average listener. On the first level, listeners react to musical changes; after a period they begin to absorb the ideas, which become important, but rarely influence the way listeners perceive an album sequence, even in a poet's album (e.g. Leonard Cohen, Bob Dylan or Joni Mitchell). Listening to music is first and foremost an emotional experience. If the musical ending of one song does not flow well into the musical beginning of the next, then the album loses its musicality. If we were dealing with lyrics without music (poetry), the intellectual order would probably be best. I feel that the intellectual point of the album will come through even if the songs are organized mainly for musical reasons — in fact it will be reinforced by the musical flow, when the album is pleasant to listen to as a unit.

Where to Start

Before putting an album in order, it's important to have its musical gestalt in mind: its sound, its feel, its ups and downs. I like to think of an album in terms of a concert. Concerts are usually organized into sets, with pauses between the sets when the artist can catch her breath, talk briefly to the audience, and prepare the audience for the mood of the next set. On an album, a set typically consists of three or four songs, but can be as short as one. Usually the space between sets is a little greater than the typical

> *"The intellectual point of an album is reinforced by its musical flow."*

space between the songs of a set, in order to establish a breather, or mood change. Sometimes there can be a long segue (crossfade) between the last song of a set and the first of the next. This basic principle applies to all kinds of music, vocal and instrumentals; it is analogous to the spacing in a classical music album: shorter ones between movements of a single composition and longer between the compositions themselves.

To make the job of organizing the sets easier, I (or the artist) prepare a rough CD of all the songs, or a DAW playlist to allow instant play of all the candidates — which is a lot easier now than it was in the days of analog tape. Then I make a simple list, describing each song's salient characteristics in one or two words or symbols, e.g. uptempo, midtempo, ballad. Sometimes I'll give letter grades to indicate which songs are the best-performed, most exciting or interesting, trying to place some of the highest grade songs early in the order for a good first impression. I may note the key of the song, although this is usually secondary compared to its mood and how it kicks off. If there's a bothersome clash in keys, sometimes more spacing helps to clear the ear, or else I exchange that song with one in a more compatible key.

The opening track is the most important; it sets the tone for the whole album and must favorably prejudice the listener. It doesn't have to be the hit or the single, but most frequently is up-tempo and establishes the excitement of the album. Even if it is an album of ballads, the first song should be the one that is most likely to engage the listener's emotions.

If the first song was up-tempo or exciting, we usually try to extend the mood, keep things moving like a concert, by a short space, with an up- or mid-tempo follow-up. Then, it's a matter of deciding when to take the audience down for a breather. Shall it be a three or four-song set? I examine the other available songs, then decide if the third song will be mid-tempo or fast followed by a relaxed fourth, or end the set with a nice, relaxed third song. Once I pick a candidate for the next song in a set, I then play the last 30 or 40 seconds of the previous song in the DAW, switching to the candidate to see if the flow is as predicted. If it works, then I pencil a checkmark in the list and move forward.

If a the transition between two songs doesn't sound good, then the sequence is faulty regardless of how compatible the bodies of the two songs seem to be. That's why transitions can let us join different musical feels; an up-tempo song that winds down gently can easily lead to a ballad. If the set doesn't flow, I try different songs until it does.

Then, I try to decide if the second set should start off with a bang or be gentle, depending on the mood that the last song of the previous set put me in. Quite often the second set also starts off with an up-tempo number in a similar "concert" pattern. This can be reversed; some sets may begin with a ballad and end with a rip-roaring number, largely depending on the ending mood from the previous set. A set can also be a roller coaster ride, if we want to create a varying mood; regardless, starting with the concept of sets makes sequencing into a "modular" job and thus a lot easier. But the ultimate listener doesn't realize there are sets in the album; our work should be subliminal. Of course there are concept albums where everyone realizes the sequence is quite special. Everyone has their favorite album transition, like in *Sergeant Pepper* between the rooster crow and Good Morning, Good Morning!

As the list gets filled up, it becomes a jigsaw puzzle to make the remaining pieces fit. Perhaps the third or fourth set doesn't work quite as well as the first, or one of the transitions is clashing, even if we increase the spacing. At that point I may try a one-song set, or try to place this problem song into an earlier set, either replacing a song, or adding to the earlier set.

The Odd Man Out

One song may just not fit well musically with the rest. For a Brazilian samba album, the artist also recorded a semi-rock blues number. She said everyone loved this song in Brazil, so we couldn't excise it from the album, but stylistically it did not gel as a part of any set. At first I suggested putting it last as a "bonus track," but this ruined the original album ending, which was a beautiful, introspective song that really did belong at the end. Eventually, we found a place for the offender near the middle of the sequence, as a one-song-set, with a long-enough pause before and after. It served as a bridge between the two halves of the album.

The Right Kind of Ending

How to end the album? What is the final encore in a concert — it's almost never a big number, because the audience always cries "more, more, more." It's better to leave them in a relaxed, comfortable "goodbye mood," or you'll be playing encores forever. That's why the last encore is usually an intimate number, or a solo, with fewer members of the band. The same principle applies to a record album. I usually try to create a climax to the entire sequence, followed by a dénouement. The climax is usually an exciting song that ends with a nice peak. Then we close out the album with one or two easy-going songs.

Once we have found the perfect sequence, it's a real treat! Rarely does this creative process take more than a few hours.

Appendix II. Radio Processing

In the 21st century, radio is undergoing serious changes: terrestrial radio is in competition with satellite and internet radio (streaming). For portable media player (PMP) fans, FM is no longer top dog. In most cases they can't receive FM, but they can listen to streaming over cellular networks or play songs located directly on their PMPs. In previous editions of this book, we concentrated on what happens to recordings when they are played over FM Radio. But now there are three possibilities:

1) *Traditional processing*, à la old-style FM radio: Audio is squished, squashed, and hyperprocessed so that everything is loud and nothing is soft. This is the culmination of many years of a radio loudness race, where no one will give an inch, or rather, a decibel for the sake of sound quality — in the name of commercial dollars.

2) *Pure Loudness normalization*: This is increasingly the case for terrestrial radio in European and other countries subscribing to the EBU's R-128 recommendation — with great success, I might add, except in the United States — where option 1 is the only kind of sound on commercial FM, except perhaps classical music stations. However, internet stations like iTunes Radio, Spotify, Pandora and many others engage in little or no processing other than adjusting the loudness to a standard target, as described in Chapters 16 and 17. I urge you to compare the best of the breed, iTunes Radio, with a high bitrate and sufficiently low target level, against its competition that is often processed or poorly normalized.

3) *A hybrid of #1 and #2*: Loudness normalization with some processing, prevalent on some internet stations and to my ears also on satellite radio. This occurs when the Internet stations engage in a loudness race of their own, raising the target for loudness normalization above a nominal -16 LUFS. Using a target higher than

about -16 LUFS either requires some peak limiting to prevent overload, or a relaxation of normalization. If normalization is reduced, some low-level programs will not be able to be brought up to the target level. The more aggressive stations appear to be taking the route of running a higher target and applying some program compression or peak limiting. This last practice gives me the most pause, because internet streaming affords the opportunity to start fresh without a loudness race, with standardized levels, bringing us a true representation of the dynamics of our masters, transmitted to the listener without sonic alteration.

The way to get the most performance out of any of the three options is not to overprocess your music. Hyper-compressed recordings are incompatible with lossy coding, especially at low bitrates, as we saw in Chapter 16, page 226. Codecs are designed to work with normally-processed music, which is not very dense. The denser the music, the greater the chance a low bitrate codec will run out of bits and distort. In satellite radio (which uses an extremely low bitrate, perhaps as low as 96 kbps), to make the recording semi-presentable on the air, engineers are likely to severely filter or lower the level of dense material.

In the case of commercial FM, the CD loudness race has been self-defeating. When you try to push music up, FM radio processing will pull it down, in the most unpleasant-sounding manner. But since every recording sounds somewhat distorted and continuously loud on commercial FM, listeners and engineers do not instantly notice that their hypercompressed recording is that much worse than a more conservatively processed recording. However, there is no escaping the transparency of pure loudness normalization where the superiority of no processing becomes intuitively obvious, especially since musicians and producers, our clients, can immediately compare the sound of their audio file on the same DAC and computer that plays the radio.

Introducing Guest Contributors Orban and Foti

Do you ever wonder what happens to your recording when it is played on the FM radio? Do you want to know how to get the most out of radio play? In 2000, participants in the Mastering Webboard engaged in a friendly collaboration to find out what range of levels we are using. Engineer Tardon Feathered offered a rock and roll mix on his FTP site, which was then mastered by webboard engineers. Tardon produced a two-CD collection of these masters called *What Is Hot?* The program loudness of the cuts on this compilation ranges from extremely hot and highly distorted to very light, with a loudness difference of 9 dB!

Next, the Webboard participants felt it would be important to demonstrate what happens to these cuts when passed through radio processing. Enter Bob Orban, who volunteered to process the music with typical radio station presets. Tardon then prepared a compilation CD, comparing the songs before and after radio processing. The figure (page 351) is a comparison of five mastered cuts before and after Orban processing, ordered from softest to loudest master, left to right.

Notice that regardless of the original level, after radio processing every source ends up with similar density: soft passages are raised radically, and loud passages are slammed to a maximum limit. Notice how the amplitude of the isolated last peak in the first example is somewhat preserved, while in all the other examples, it is pushed down compared to the body of the song, giving rise to a "dynamic reversal" effect, which is very common on the radio. Since the first master is lower in level than the others, it demonstrates that if you push the signal up, the radio will pull it back down, in an unpleasant manner. Listening to this revealing CD, every track ends up at the same loudness, proving beyond a shadow of a doubt that there is no advantage to "compressing for the radio." In addition, the radio processing yields negative effects from extremely compressed tracks, distorting nearly every original, except for the softest track, which came in at about -17 LUFS. The rightmost and most squashed source track was unlistenable after Orban processing. The radio processing also somewhat randomizes the stereo image and lowers the high end, but attempting to compensate for these losses in mastering only aggravates the distortion and does not help the clarity of the final product.

Please meet guest contributors — Bob Orban and Frank Foti.[1] They are considered to be the world's authorities on radio processing. Bob is the engineer and designer of the Optimod line of audio processors, while Frank is the creator and lead designer of the Omnia product line. Together, their products are used by nearly every FM radio station around the world. Here's what Orban and Foti say about what goes on inside the box...

What Happens to My Recording When it's Played on the FM Radio?
by Robert Orban and Frank Foti[2]

Few people in the record industry really know how a radio station processes their material before it hits the FM airwaves. This article serves to remove the many myths and misconceptions.

Every radio station uses a transmission audio processor in front of its transmitter. The processor's most important function is to control peak modulation to the legal requirements of the regulatory body in each station's nation. However, very few stations use a simple peak limiter for this function. Instead, they use more complex audio chains. These can accurately constrain peak modulation while significantly decreasing the peak-to-average ratio of the audio. This makes the station sound louder within the allowable peak modulation.

> *"The main obstacle to good digital radio sound is low bit rate."*

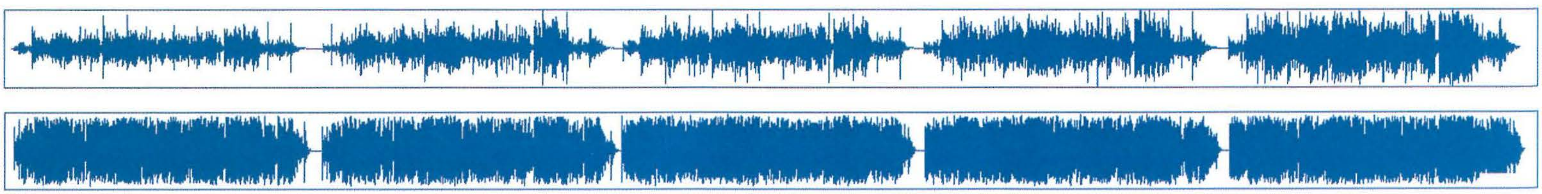
Garbage In — Garbage Out

Manufacturers have tuned broadcast processors to process the clean, dynamic program material that the recording industry has typically released throughout its history. (The only significant exception that comes to mind is 45 rpm singles, which often were overtly distorted.) Because these processors have to deal with speech, commercials, and oldies in addition to current material, they can't be tuned exclusively for "hypercompressed," distorted CDs. Indeed, experience has shown that there's no way to tune them successfully for material which has arrived so degraded.

For 20 years, broadcast processor designers have known achieving highest loudness consistent with maximum punch and cleanliness requires extremely clean source material. Orban's published application notes to help broadcasters clean up their signal paths. These notes emphasize that clipping in the path before the processor causes subtle degradation that the processor will often exaggerate severely. They promote adequate headroom and low-distortion amplification to prevent clipping even when an operator drives the meters into the red.

Top, five mastered cuts of the same music, with increasing loudness and visual density.

Bottom, the same cuts passed through the Orban radio processor.



About 1997, we started to notice CDs arriving at radio stations that had been pre-distorted in production or mastering to increase their loudness. For the first time, we started seeing frequently recurring flat topping caused by brute-force clipping in the production process. Broadcast processors react to pre-distorted CDs exactly the same way as they have reacted to accidentally clipped material for more than 20 years — they exaggerate the distortion. Because of phase rotation, the source clipping never increases on-air loudness — it just adds grunge.

The authors understand the reasoning behind the CD loudness wars. Just as radio stations wish to offer the loudest signal on the dial, it is evident that recording artists, producers, and even some record labels want to have a loud product that stands out against its competition in a CD changer or a music store's listening station.

In radio broadcasting this competition has existed since about 1975,[3] when radio stations used simple clipping to get louder, and this technique has now migrated to the music industry. The figure (page 353) shows a section of a severely clipped waveform from a contemporary CD.

The area marked between the two pointers highlights the clipped portion. This is one of the roots of the problem; the other is excessive digital limiting in CDs that does not necessarily cause flat-topping, but still removes transient punch and impact from the sound.

The problem today is that sophisticated and powerful audio processing for the broadcast transmission system does not coexist well with a signal that has already been severely clipped. Unfortunately, with current pop CDs, the example shown is more the norm than the exception.

The attack and release characteristics of broadcast multiband compression were tuned to sound natural with source material having short-term peak-to-average ratios typical of vinyl or pre-1990 CDs. Excessive digital limiting of the source material radically reduces this short-term peak-to-average ratio and presents the broadcast processor with a new, synthetic type of source that the broadcast processor handles less gracefully and naturally than it handles older material. Instead of being punchy, the on-air sound produced from these hypercompressed sources is small and flat, without the dynamic contours that give music its dramatic impact. The on-air sound resembles musical wallpaper and makes the listener want to turn down the volume control to background levels.

There is a myth that broadcast processing will affect hypercompressed and clipped material less than it will more naturally produced material. This is true in only one aspect — if there is no long-term dynamic range coming in, then the broadcast processor's AGC[4] will not further reduce it. However, the broadcast processor will still operate on the short-term envelopes of hypercompressed material and will further reduce the peak-to-average ratio, degrading the sound even more.

Hypercompressed material does not sound louder on the air. It sounds more distorted, making the radio sound broken in extreme cases. It sounds small, busy, and flat. It does not feel good to the listener when turned up, so he or she hears it as background music. Hypercompression, when combined with "major-market" levels of broadcast processing, sucks the drama and life from music. In more extreme cases, it sounds overtly distorted and is likely to cause tune-outs by adults, particularly women.

A Typical Processing Chain—What Really Goes On When Your Recording is Broadcast

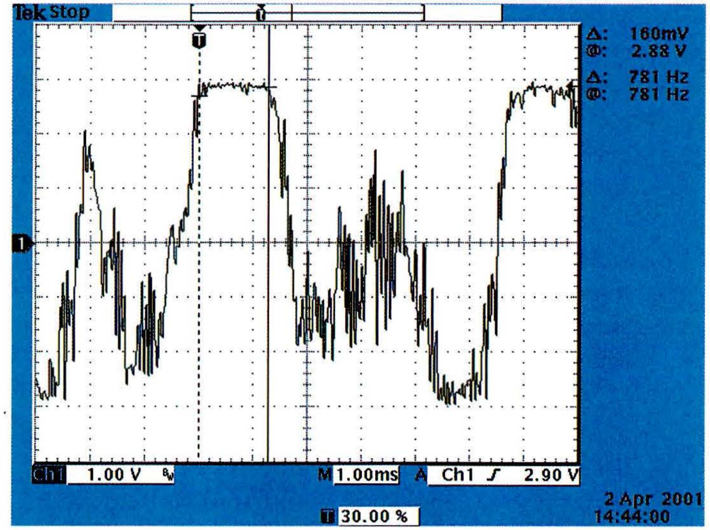
A typical chain consists of the following elements, in this order:

Phase rotator. The phase rotator is a chain of allpass filters (typically four poles, all at 200 Hz) whose group delay is very non-constant as a function of frequency. Many voice waveforms (particularly male voices) exhibit as much as 6 dB asymmetry. The phase rotator makes voice waveforms more symmetrical and can sometimes reduce the peak-to-average ratio of voice by 3-4 dB. Because this processing is linear (it adds no new frequencies to the spectrum, so it doesn't sound raspy or fuzzy) it's the closest thing to a "free lunch" that one gets in the world of transmission processing.

There are a few prices to pay. In the good old days when source material wasn't grossly clipped, the main price was a very subtle reduction in transparency and definition in music. This was widely accepted as a valid trade-off to achieve greatly reduced speech distortion, because the phase rotator's effects on music are unlikely to be heard on typical consumer radios, like car radios, boomboxes, "Walkman"-style portables, and table radios.

However, with the rise of the clipped CD, things have changed. The phase rotator radically changes the shape of its input waveform without changing its frequency balance: If you measured the frequency response of the phase rotator, it would measure "flat" unless you also measured phase response, in which case you would say that the "magnitude response" was flat and the phase response was highly non-linear with frequency. The practical effect of this non-linear phase response is that flat tops in the original signal can end up anywhere in the waveform after processing. It's common to see them go right through a zero crossing. They end up looking like little smooth sections of the waveform where all the detail is missing — a bit like a scar from a severe burn. This is an apt metaphor for their audible effect, because they no longer help reduce the peak-to-average ratio of the waveform. Instead, their only effect is to add unnecessary grungy distortion. Thanks to phase rotation, any clipping in the source material causes nothing but added distortion without increasing on-air loudness at all.

AGC. The next stage is usually an average-responding AGC. By recording studio standards, this AGC is required to operate over a very wide dynamic range — typically in the range of 25 dB. Its function is to compensate for



A severely-clipped waveform from a contemporary CD

operator errors (in live production environments) and for varying average levels (in automated environments). Average levels vary mainly because the peak to average ratio of CDs themselves has varied so much from about 1990 on. Peak-normalizing hard disk recordings (to use all available headroom) has the undesirable side effect of causing gross variations in average levels. Indeed, 1:1 transfers (which are also common) will also exhibit this variation, which can be as large as 15 dB![5]

The price to be paid is simple: the AGC will eliminate long-term dynamics in your recording. Virtually all radio station program directors want their stations to stay loud always, eliminating the risk that someone tuning the radio to their station will either miss the station completely or will think that it's weak and can't be received satisfactorily. Radio people often call this effect "dropping off the dial."

AGCs can be either single band or multiband. If they are multiband, it's rare to use more than two bands because AGCs operate slowly, so "spectral gain intermodulation" (such as bass pumping the midrange) is not as big a potential problem as it is for later compression stages, which operate more quickly.

AGCs are always gated in competent processors. This means that their gain essentially freezes if the input drops below a preset threshold, preventing noise suck-up despite the large amount of gain reduction.

Stereo Enhancement. Not all processors implement stereo enhancement, and those that do may implement it somewhere other than after the AGC. (In fact, standalone stereo enhancers are often placed in the program line in front of the transmission processor.)

The common purpose of stereo enhancement is to make the signal stand out dramatically when the car radio listener punches the tuning button. It's a technique to make the sound bigger and more dramatic. Overdone, it can remix the recording. Assuming that stereo reverb, with considerable L–R energy, was used in the original mix, stereo enhancement, for example, can change the amount of reverb applied to a center-channel vocalist. The moral? When mixing for broadcast, err on the "dry" side, because some stations' processors will bring the reverb more to the foreground.[6]

Because each manufacturer uses a different technique for stereo enhancement, it's impossible to generalize about it. The only universal constraints are the need for strict mono compatibility (because FM radio is frequently received in mono, even on "stereo" radios, due to signal-quality-trigged mono blend circuitry), and the requirement that the stereo difference signal (L–R) not be enhanced excessively. Excessive enhancement always increases multipath distortion (because the part of the FM stereo signal that carries the L–R information is more vulnerable to multipath). Excessive enhancement will also reduce the loudness of the transmission (because of the "interleaving" properties of the FM stereo composite waveform, which we won't further discuss).

These constraints mean that recording-studio-style stereo enhancement is often incompatible with FM broadcast, particularly if it significantly increases average L–R levels. In the days of vinyl, a similar constraint existed because of the need to prevent the cutter head from lifting off the lacquer, but with CDs, this constraint no longer exists. Nevertheless, any mix intended for airplay will yield the lowest distortion and highest loudness at the receiver if its L–R/L+R ratio is low. Ironically, mono is loudest and cleanest!

Equalization. Equalization may be as simple as a fixed-frequency bass boost, or as complex as a multistage parametric equalizer. EQ has two purposes in a broadcast processor. The first is to establish a signature for a given radio station that brands the station by creating a "house sound." The second purpose is to compensate for the frequency contouring caused by the subsequent multiband dynamics processing and high frequency limiting. These may create an overall spectral coloration that can be corrected or augmented by carefully chosen fixed EQ before the multiband dynamics stages.

Multiband Compression and Limiting. Depending on the manufacturer, this may occur in one or two stages. If it occurs in two stages, the multiband compressor and limiter can have different crossovers and even different numbers of bands. If it occurs in one stage, the compressor and limiter functions can "talk" to each other, optimizing their interaction. Both design approaches can yield good sound and each has its own set of tradeoffs.

Usually using anywhere between four and six bands, the multiband compressor/limiter reduces dynamic range and increases audio density to achieve competitive loudness and dial impact. It's common for each band to be gated at low levels to prevent noise rush-up, and manufacturers often have proprietary algorithms for doing this while minimizing the audible side effects of the gating.

An advanced processor may have dozens of setup controls to tune just the multiband compressor/limiter. Drive and output gain controls for the various compressors, attack and release time controls, thresholds, and sometimes crossover frequencies are adjustable, depending on the processor design. Each of these controls has its own effect on the sound, and an operator needs extensive experience if he or she is to tune a broadcast multiband compressor so that it sounds good on a wide variety of program material without constant readjustment. In broadcast there's no mastering engineer available to optimize the processing for each new source!

Pre-Emphasis and HF Limiting. FM radio is pre-emphasized at 50 microseconds or 75 microseconds, depending on the country. Pre-emphasis is a 6 dB/octave high frequency boost that's 3 dB up at 2.1 kHz (75 µs) or 3.2 kHz (50 µs). With 75 µs pre-emphasis, 15 kHz is up 17 dB!

Depending on the processor's manufacturer, pre-emphasis may be applied before or after the multiband compressor/limiter. The important thing for mixers and mastering engineers to understand is that putting lots of energy above 5 kHz creates significant problems for any broadcast processor because the pre-emphasis will greatly increase this energy. To prevent loudness loss, the processor applies high-frequency limiting to these boosted high frequencies. HF limiting may cause the sound to become dull, distorted, or both, in various combinations. One of the most important differences between competing processors is how effectively a given processor performs HF limiting to minimize audible side effects. In state-of-the-art processors, HF limiting is usually performed partially by HF gain reduction and partially by distortion-cancelled clipping.

Clipping. In most processors, the clipping stage is the primary means of peak limiting. It's crucial to broadcast processor performance. Because of the FM pre-emphasis, simple clipping doesn't work well at all. It produces difference-frequency IM distortion, which the de-emphasis in the radio then exaggerates. (The de-emphasis is flat below 2-3 kHz, but rolls off at 6 dB/octave thereafter, effectively exaggerating energy below 2-3 kHz.) The result is particularly offensive on cymbals and sibilance ("esses" become "efffs").

In the late seventies, Bob Orban invented distortion-cancelled clipping. This manipulates the distortion spectrum added by the clipper's action. In FM, it typically removes the clipper-induced distortion below 2 kHz (the flat part of the receiver's frequency response). This typically adds about 1 dB to the peak level emerging from the clipper, but, in exchange, allows the clipper to be driven much harder than would otherwise be possible.

Provided that it doesn't introduce audibly offensive distortion, distortion-cancelled clipping is a very effective means of peak limiting because it affects only the peaks that actually exceed the clipping threshold and not surrounding material. Accordingly, clipping does not cause pumping, which gain reduction can do, particularly when gain reduction operates on pre-emphasized material. Clipping also causes minimal HF loss by comparison to HF limiting that uses gain reduction. For these reasons, most FM broadcast processors use the maximum practical amount of clipping that's consistent with acceptably low audible distortion.

Real-world clipping systems can get very complicated because of the requirement to strictly band-limit the clipped signal to less than 19 kHz despite the harmonics that clipping adds to the signal. (Bandlimiting prevents aliasing between the stereo main and subchannel, protects subcarriers located above 55 kHz in the FM stereo composite baseband, and protects the stereo pilot tone at 19 kHz). Linearly filtering the clipped signal to remove energy above 15 kHz causes large overshoots (up to 6 dB in worst case) because of a combination of spectral truncation and time dispersion in the filter. Even a phase-linear lowpass filter (practical only in DSP realizations) causes up to 2 dB overshoot. Therefore, state-of-the-art processors use complex overshoot compensation schemes to reduce peaks without significantly adding out-of-band spectrum.

Some chains also apply composite clipping or limiting to the output of the stereo encoder, which encodes the left and right channels into the multiplex signal that drives the transmitter. It's actually the peak level of this signal that government broadcasting authorities regulate. Composite clipping or limiting has long been a controversial technique, but the latest generation of composite clippers or limiters has greatly reduced interference problems characteristic of earlier technology.

Conclusions

Broadcast processing is complex and sophisticated, and was tuned for the recordings produced using practices typical of the recording industry during almost all of its history. In this historical context, hypercompression is a short-term anomaly and does not coexist well with the "competitive" processing that most pop-music radio stations use. We therefore recommend that record companies provide broadcasters with radio mixes. These can have all of the equalization, slow compression, and other effects that producers and mastering engineers use artistically to achieve a desired "sound." What these radio mixes should not have is fast digital limiting and clipping. Leave the short-term envelopes unsquashed. Let the broadcast processor do its work. The result will be just as loud on-air as hypercompressed material, but will have far more punch, clarity, and life.

A second recommendation to the record industry is to employ studio or mastering processing that provides the desired sonic effect, but without the undesired extreme distortion from clipping. The alternative to brute-force clipping is digital look-ahead limiting, which is already widely available to the recording industry from a number of different manufacturers (including the authors' companies). This processing creates lower modulation distortion and avoids blatant flat-topping of waveforms, so is substantially more compatible with broadcast processing. Nevertheless, even digital limiting can have a deleterious effect on sound quality by reducing the peak-to-average ratio of the signal to the point where the broadcast processor responds to it in an unnatural way, so it should be used conservatively. Ultimately, the only way to tell how one's production processing will interact with a broadcast processor is to actually apply the processed signal to a real-world broadcast processor and to listen to its output, preferably through a typical consumer radio.

1
Robert Orban, Orban Inc. (A CRL Company). Frank Foti, Omnia Audio

2
Edited and adapted from a 2001 AES presentation.

3
Bob Ludwig (in correspondence) mentions that competition in radio broadcasting was already happening in the late 1960's, noting WABC "color radio" added EMT plate to everything to increase average density.

4
Automatic Gain Control. A compressor that brings up low-level passages. See Chapter 7.

5
No wonder CD changers are a predicament. See Chapter 17.

6
On the other hand, the other radio processing, especially the compression, reduces depth, plus, distant reception areas tend to lose separation so I'm not so sure that improving the stereo image in mastering is such a bad thing. My approach is to make a recording sound good at home first, then let the radio translate it as well as it can.

Appendix III. Preparing Tapes and Files for Mastering

One major theme in this book has been the mastering engineer's comprehensive attention to sequencing, spacing (a.k.a. assembly), leveling, clean-up and processing. The better-prepared the mix tape or file, the better we all will look. Make the best mix you can, then let the mastering engineer do the rest of the magic, including the "heads, tails, fade-ins and fadeouts," for if something is cut off or faded prematurely, it will be lost. Don't be tempted to fade even if there is a noise, because we can create natural-sounding endings on tunes that everyone thought had to be faded, as described in Chapter 3. If you must fade, then also include an unfaded ending. We can then either use the faded version or treat it as a "fade example," if we can do it better in our DAWs. Given freedom, the mastering engineer can produce a seamless, flowing record album from the "loose parts" sent by the mix engineer. Leaving the tunes loose also permits the mastering house the most flexibility to change the order of the album (if necessary), or produce segues in the most artistic fashion.

In the last century, the most common formats we received for music mastering were linear, e.g. analog and digital tape. But now the most popular formats are completely random access (file-based). Here's how to satisfy the needs of the mastering engineer when submitting finished mixes on the medium of your choice.

Communication

Mastering is a collaborative process, even if you cannot attend the mastering session; the mastering engineer's job is to realize your desires and if possible to go beyond your wildest dreams! The mix engineer or producer should discuss the music and his goals and involve the mastering engineer at an early stage; if in the neighborhood, bring over a sample to hear on the high-resolution, wide-range mastering monitors. Listen: Does it sound like music? Does it live and breathe? Do the climaxes sound like climaxes? Do the choruses have a bit more energy than the verses (as they naturally should)? Is the bass drum to bass ratio right? Is the sound as spacious and deep as you want it to be? How does it sound on several alternative systems? When the mastering date arrives, don't hesitate to provide or suggest a recording of similar music that appeals to you, yet leave your mind open to the creativity of the mastering engineer. When the mastering session is over, the mastering engineer will provide a reference disc or file that the producer will check and if desired, suggest revisions or improvements.

Mix Logs

The logs that accompany mixes are very important. Logs keep a project from being delayed because we don't have to chase down the catalog number or other essential information on the mastering day. Some engineers forget that a disc of files has no order.[1] So all logs should indicate the full title of each song, the corresponding file name on disc, and where the song is to appear on the final medium, plus comments about fades, noises, any problem or concern, or special requests. ("Please leave that ugly laugh in between songs 2 and 3, I think it's funny.") Appendix V contains example logs.

Stems, Splits and Alternate Mixes (e.g. Vocal Up/Down).

Send an example mix to the mastering engineer long before the mixing is complete. If he suggests stems or alternate mixes for technical reasons relating to your mix, then include these if possible. See Chapter 9.

Linear Media (Analog tape)

Digital audio tape machines are pretty much obsolete, so the last extant linear medium is analog tape. Don't bother to reorder analog tapes. Leave the tunes out of order, leave the outtakes and alternate mixes (which may prove useful), and mark all keeper takes in the log. Don't bother to space the tunes on linear media other than leaving enough time to cue and to use leaders to identify the cuts. If possible simultaneously mix to analog tape and a digital medium as a safety and send that along. Make a digital safety of the analog tape at 2496 or higher rate and keep that in case the tape gets damaged or lost.

When mixing to a linear digital recorder, always record two at once (make data-identical mixes labeled "A" and "B"), and hold onto that safety — never send the only copy in the mail. When mixing to a hard disk, make sure you have backups (mirrors) of every critical hard disk. Even **RAID**s should be backed up to another medium.

Level Check

Avoid fast digital peak limiters in mixing and mastering [I wish I could follow my own advice!], as described in Chapter 16. As Orban and Foti pointed out in Appendix II, fast digital limiters are bad for subsequent radio processing. In a perfect world, we should also avoid fast digital limiting in mastering, except for rare esthetic purposes. As described in Chapter 16, mix with conservative levels, which is not a problem with 24-bit media. Print the mix with peak levels well under the top and no OVERs! I recommend -3 dBFS maximum or -1 dBFS maximum if you have accurate metering. There's nothing wrong with a 24-bit mix that never exceeds -10 dBFS peak. Even if working in floating point, it may be advisable to ensure that each plugin does not overload, unless you are certain that your plugins behave in a classic floating point manner, as described in Chapter 16.

Preserve Data Integrity

In general, send the earliest generation, unprocessed material to the mastering house. If you must edit, keep everything at unity gain if at all possible (do not normalize), even if the material is peaking low, as explained in Chapter 16. The same goes for temptation to equalize, compress, limit or process a mix after it has been made. If you do post-process because you discover something whose sound you love, please send both versions to the mastering house. We may have a better approach with our tools, or we may discover that combining your process with ours produces an unexpected result. Communicate, send test mixes to ensure that all questions will be settled long before the mastering date.

Maximum CD Program Length

Every replication plant specifies a maximum acceptable length, and some charge more for CDs over approximately 77 minutes. The final master, including songs, spaces between songs, and reverberant decay at the ends of songs, must not exceed the limit, which at one popular plant is 79:38. But don't take that as gospel: check with the plant that is replicating the CD. The mastering engineer can determine the exact time after the master is assembled. DVD program lengths vary because of the data coding and must be determined at the time of authoring.

Labeling tapes or discs. Which is the Master?

Don't forget to put a name and phone number on the source media in case it gets separated from the documentation! The sources for an album are NOT the master: the album (production) MASTER is the final, PQ'd, equalized, edited, assembled, and prepared tape, disc or file that needs no further audio work, and is ready for replication.[2] Please label the source media: Submaster, Work Tape, Mix, Final Mix, Session Tape, Edited-Mix, Compiled-mix, or Equalized Mix, to name several possibilities. This will avoid confusion in the future when looking through the tape library for the one and only real (production) master.[3]

Analog tape Preparation

Begin and end the reel with some "bumper," followed by leader. If possible, put leader between songs (except for live concerts and recordings edited with room tone). Tape should be slow wound, tail out. Label each reel as recommended in Appendix V. Indicate tape speed, record level for 0 VU in nw/m, record EQ (NAB or IEC), track configuration, whether it is mono, stereo or multichannel. Indicate if noise reduction is used and include the noise reduction alignment tone. Include alignment tones 30 seconds (or longer) each, at 0 VU, without noise reduction, minimum 1 kHz, 10 kHz, 15 kHz, and 100 Hz plus (highly recommended) 45 Hz and 5 kHz. Also highly recommended is a tone sweep (glide) from 20 Hz through 500 Hz. The tones must be recorded by the same tape recorder that recorded the music, and ideally, recorded through the same console and cables that were used to make the mix. Many mastering engineers prefer having the tones at the tail of a reel or on a separate reel. Store all tapes flat wound, tail out.

Many historic analog tapes do not include proper tones and sometimes it is not possible to put tones on new masters. If it was not possible to lay down tones on the session, then we will use sophisticated methods to guarantee azimuth and equalization accuracy.

Include Handles

As described in Chapter 3, for live concerts and many other forms of music, it's useful to include handles, that is, raw footage on either side of the intended music. This can include outtakes, unfaded applause, breaths, coughs, noises, speech between tracks, etc. Handles are especially useful when a track might have to be noise reduced, for the noise sample we need can sometimes only be found just before the downbeat.

What Sample Rate?

Until around 2000, I recommended that mix engineers try to work at 44.1 kHz if possible for CD, considering the then poor state of typical sample rate converters. This is no longer necessary nor desirable; high quality sample rate converters can convert between 96 kHz and 44.1 kHz with high integrity, as described in Chapter 23. The best recommendations are for mix engineers to work at the highest practical sample rate and longest available wordlength. However, there is some reasonable opinion on whether double sample rates are not as valid for rock and roll because higher fidelity does not suit the genre, but regardless, I think it is a good idea to record and mix at least at 2448 so as to avoid cumulative losses after the mix stage.

"The source tapes/files for an album are NOT the Production MASTER."

If you are mixing digitally, do not sample-rate convert yourselves, but remain at the same sample rate as the multitrack. If you are mixing with an analog console, there is a marginal advantage to using a higher sample rate for the mixdown recorder than the multitrack. For example, even if mixing analog with a multitrack at 48 kHz, I think you will get slightly better results with a mixdown medium at 96 kHz, especially if you are pushing levels hard (See Chapter 22).

Preparing Files

Whether you send an optical disc with files on it or send files over FTP to the mastering house, files require attention to detail, as poorly-prepared files can waste a tremendous amount of time at the mastering house. Make sure the mastering house will accept the file types you want to send. Here are some critical do's:

- Leave blank sound at the head of the file: in other words, start the first music at least 1 second into the file, not at zero time (this is to prevent glitches that often occur at the file start).
- For stereo and multichannel, high-resolution, linear-format, interleaved files are preferred. Mono or split stereo files are also acceptable. WAV or AIFF are functionally equivalent, either format is acceptable. No mp3's or AACs, please!
- Try to do one project at a single sample rate. However, if

for some reason your project includes different rates, do not convert them yourself. Instead, carefully mark (log) the rate of the files for our information. In mastering I upsample all material to a common sample rate, usuall 88.2 kHz or 96 kHz, but some mastering engineers work at the native rate, and files coming in at different rates in the same project can be problematic.

- Give each file a meaningful name related to the song title, like *Love_Me_Do.wav*, not some meaningless serial number.
- If burning an optical disc, choose a high-quality name-brand optical blank. To my experience, Taiyo Yuden, the oldest manufacturer, continues to make the most compatible and reliable blank CDRs.
- Write a Fixed optical disc, i.e. a closed session.
- Track IDs on CD-DA mixes do not have to be exactly placed, but they guide us to loading the proper tune according to your log. But since 24-bit and 32-bit float mixes are far more preferable to 16-bit, we don't see many CD-DA sources anymore.
- DO NOT USE PAPER LABELS! Stick-on paper labels may look impressive, but they increase error rate and they are dangerous at high rotational speeds. Labels have become partly or completely unglued and tear off in the reader, which is not a pretty sight! Also, do not label the disc with a ball-point pen, but with a soft marker, on the protected (overcoated) part of the top surface.[4] While I personally believe that the coating on professionally over-coated CDRs is sufficient protection from scratches and organic solvents (as in an aromatic Sharpie-brand marker), the most conservative mastering engineers recommend using water-based markers for labeling. Perhaps someone will do a long-term study measuring errors on CDRs with a coated-marked surface. Please do not send Sharpies over the Internet, they mess with the bits.
- Find out if the mastering house can accept 32-bit floating point files, and in which of several competing 32-bit formats. The most commonly used is compatible with Nuendo and Samplitude. When in doubt, write to fixed-point 24-bit files (also known as Integer Format). Don't forget to use dithering to 24 bits when creating 24-bit files (See Chapter 15).
- Use any standard sample rate up to 96 kHz. Verify the mastering house can use files with a higher rate before cutting.
- Do not use the / or \ character in a filename. The file could disappear once it has been copied to another operating system! For best multi-platform compatibility, use only alphas and numerics. No curly quotes, no accent marks.[5]
- I love receiving files that include the intended track number in their name. It is very useful to include the intended track number at the beginning of the file name (using two digits), which makes it much easier to assemble them in the album order. For example: 01 I Need Somebody, 09 I Got Rhythm, 10 She's So Fine.

Split Files

Interleaved files are less subject to accidents since all the channels are guaranteed to start at the same point. For multichannel, include a note indicating the channel order used, e.g., L, R, C, LFE, SL, SR or L, R, SL, SR, C, LFE. If you must send split files, use a standard nomenclature to distinguish the channels, e.g. Do It.L.wav, Do It.R.wav, Do It.SL.wav, Do It.SR.wav, Do It.C.wav, Do It.LFE.wav. Letter abbreviations are preferable to ambiguous channel numbers.

When You Get Your Master Back

If the mastering house is producing a physical CDR master and the master is sent back to you instead of directly to the CD plant, don't handle it or play it. Play the ref, not the master!

1
There is no track order on a non-linear, file-based medium. Often, clients ask me to "put the master in the order that's on the CD-ROM," but they forget the only order on the CD-ROM is the alphabetical directory of files.

2
Andre Subotin on the Mastering Webboard reminds us that there may be several true Masters, each of which we must clearly label, e.g. Production Master for Cassette; Master for foreign countries; etc.

3
Visit the NARAS Master Recording Delivery Recommendations, in the links.

4
Thanks to Clete Baker and Mike McMillan on the Mastering Webboard for clarifications on these points.

5
Thanks to Clete Baker on the Mastering Webboard for reminding me of this essential!

Appendix IV. Premastering for Vinyl

To make the best premaster for vinyl, do nothing special, work with conservative levels and good headroom because the level of the premaster has no effect on the level of the vinyl cut. The vinyl cutting engineer decides the level he can cut based on several factors, including the amount of bass information and the length of the side. A little better sound quality on vinyl can be obtained by creating a data disk with 24-bit files rather than a 16-bit audio CD. I tend to make 2496 files as masters for vinyl. We include the desired space at the tail of each file and he puts them in contiguous order in a DAW and then outputs to the cutter, or I send a continuous wav file for each side along with a list of songs and their lengths. Duration is the only thing to be concerned about, especially when there is a lot of bass on the record. A ten-minute side is usually no problem even when there is heavy bass. It's technically possible to put half an hour on an LP side, but almost inevitably with loss of level, stereo separation and/or bass. The longest practical side length in pop music today is around 17 minutes.

Mastering engineers should not try to apply a process specifically intended to help the LP cut. When mastering for vinyl, our job is to get the *sound* we want to hear onto the source medium. Let the cutting engineer expert try to *translate* our sound to the LP medium. He may have to narrow the separation at the bass end to protect the groove excursion, and he may have to insert some high-frequency limiting to protect the cutterhead, but the latter can be good for sibilance control. Let him do what he has to for the idiosyncrasies of the medium. If there is time, you could send a sample song to him for evaluation, but there never seems to be enough time for that. Sad because I'm sure the LP cutting engineer would have a lot to contribute to the success of any project.

Appendix V. Logs and Labels for Tapes, Discs and Boxes

Logs

When uploading files via FTP, include a readme file with a log of information about the material that you are sending. The example logs below are suitable for uploaded files or printed logs to accompany optical discs sent to the mastering house.

Labeling Source Media

I don't dare put an unlabeled disc down on my mastering desk, for it will immediately be lost in a crowd! Please put the following minimal information on every piece of source media, in case it gets separated from the box:

- Artist
- Album Title [or working title]
- Contact Name, phone number
- Tape or reel number
- Date [important to help separate out revisions]

Labeling Those Boxes

The box label contains much more information than what's written on the reel or disc itself.

Analog Tape Boxes: An example label

Some studios have preprinted labels with checkboxes for each option.

Mix tape, Unedited, songs head leadered [or other descriptive]

Artist:______________________________

Album Title: __________________________

Record Label: _________________________

Reel number: ________ of_______________

Catalog Number: _______________________

Studio, Address, Contact Phone #:____________

Engineer: _____________________________

Assistant:_____________________________

Producer: ____________________________

Date: ________________________________

Format, EQ, Speed, Level: [e.g. 1/2" 2-track AES stereo, no noise reduction, 30 IPS, 0 VU = 320 nW/M, or 0 VU = 250 nW/M + 2 dB] ________________

Test Tones @ Head ______ @ Tail ______ consisting of ________________________ Hz at 0 VU

Name of Song
or Track Length ________________________

Comments [e.g. "vocal up" or "needs fadeout" or "leave countoff at the beginning" ____________

Name of next song, etc. ____________________

Further comments can be written in a letter that accompanies all the media.

Discs: Example Label

There is not enough room on a CDR or DVD-R surface to write everything we want to know. Some studios have prescreened discs with checkboxes. At minimum, the top surface of the disc itself should include:

> **Mixes, Unedited [or submaster or other descriptive]**
>
> Artist: ______________________________
>
> Album title: ______________________________
>
> Record Label: ______________________________
>
> Disc and File Format: [e.g. ISO-9660, HFS, Stereo AIFF Files, 48 kHz/24 Bit, etc.] ______________________
>
> Disc # _____ of _______ Date: _________________
>
> [date is very critical]
>
> Plus, if possible:
>
> Contact name and Phone #: ______________________
>
> Catalog Number: ______________________________

Since there is not enough room to list all information on the disc itself, be sure to include the remaining information on the box, jewel box, and/or printed log (pictured opposite page) that accompanies the media. If possible, the log can be duplicated in a READ_ME.doc file which resides on the disc, so it will never be lost.

Discs, Jewel Box or Paper cover label

Instead of using up several jewel boxes, some studios cleverly put CDRs inside a taped and folded printout of the disc's directory, which covers all the names of the tunes inside the disc. When shipping, put these paper-covered discs in a foam-lined hard-box to prevent scratching or breakage.

Log/letter

Accompany the discs, tapes, or file upload with a printed log/letter or readme file for the mastering engineer. This is where you can also include all your comments and thoughts on the eventual mastering. You can add your story and feelings about the album and its sound. Some comments may be superfluous, but put down anything you are concerned about. On page 364 is an example mix engineer's log with information for each tune.

Don't forget to include in the cover letter:

> Artist: ______________________
>
> Album title: _________________
>
> Record Label: _______________
>
> Disc, File Format, Sample Rate, Wordlength: [e.g. ISO-9660 or HFS, or Masterlink, Stereo AIFF Files, 48 kHz/24 Bit] __________
>
> Contact name and Phone #: _____
>
> Contact Address: _____________
>
> Catalog Number: _____________

Title/File Name	Track Order	Length (approx)	ISRC	Comments [e.g. by engineer, producer, artist]
I Wanna Make You Happy/ 05_makehappy.wav	5	4:02	ES60801332805	This song needs a fadeout. Try starting ~3:45 and be out by 4:00 from the downbeat so as not to hear the snickering! Please include the sticks at the beginning.
Love Me Do/ 02_lovemedo.wav	2	2:55	ES6080132802	This is an obvious tribute to the Beatles. The more Beatle-like you can make the mastering, the happier I will be.
Why Me?/ 04_yme.wav	4	5:02	ES6080132804	This is the only ballad on the album. The artist is not happy with her intonation entering from the last chorus. Is there anything you can do about this?

Mix engineer's log, with notes for fades and spacing. Logs can also be in the form of a letter to the mastering engineer which include this information.

Appendix VI. I Feel The Need for Capacity

Year	Type of Storage	Capacity GB	Total Cost U.S. Dollars	Cost per GB	# Hours/2 track/2496	Facts
1980	Data General	0.297	$35,000	$117,845	9 min.	Size: 2 feet x 3 feet x 3-1/2 feet high!
1990	SCSI Hard Disk	0.58	$750	$1,293	18 min.	Each drive specified as "600 MB." CD one hour 635,040,000 bytes
2002	IDE Hard Disk	80	$137	$1.71	41 Hr.	Street price
2014	1 TB ESATA	1024	$65	$0.06	530 Hr.	
2014	3 TB ESATA	3072	$125	$0.04	1590 Hr.	
2014	6 TB ESATA	6144	$369	$0.06	3180 Hr.	

GB = Gigabytes, based on 1 GB = 1024 MB, 1 TB = 1024 GB.

Between 1982 and 1990, I was producing CD masters with linear editing using the Sony 3/4" editing systems based on videotape transports. It took 5 to 10 minutes to rewind an 80 minute tape! Today we have instant access to any point in a project. From about 1976 until about 1988, only one computer-based hard disk editing system was available in New York City, the Soundstream at BMG with two giant Braegen 14" drives, whose capacity I have not been able to uncover. Some Braegens of that era were capable of up to 850 MB each. Around 1989 Sonic Systems became available. In 1990 I set up my first nonlinear mastering work-station, Sonic Solutions, purchasing the highest capacity hard discs available, a pair of 600 MB SCSI hard discs, that cost $1500 retail in 1990 dollars, $1.29 per MB. That's Megabyte, not Gigabyte! (Described in the second line of the above table). I had to archive the disks to DAT and then wipe them several times a week as they could only support one or two projects. Fortunately, as our needs have gone up, capacity has increased and cost has gone down exponentially. At Digital Domain we now have four different RAID arrays totaling 40 TB (that's Terabytes) of storage. Seagate promises to have 60 terabyte *individual* 3.5 inch hard drives by 2024, but I believe it will be much sooner!

Appendix VII. I Feel The Need for Speed

Medium	Theoretical Speed MB/sec	Practical Speed MB/sec	Theoretical Speed Mb/sec	Practical Speed Mb/sec	Equivalent Stereo 2496 Tracks Real time**	Notes*
CD Player (1X Speed)	0.17	N.A.	1.35	N.A.	0.3	Speed of CD player could only deliver 0.3 stereo 2496 tracks
1000 Base T Ethernet	125	56	1000	448	101	Gigabit Ethernet
USB 2.0	60	29.5 our timing (Macworld claims 41)	480	236	53	This implies USB 2.0 could deliver 53 simultaneous stereo 249 tracks, but have not actually tested
USB 3.0 (MacWorld 7200 RPM drive)	640	112	5120	896	203	Claimed "10 times faster than USB 2.0." However, the drive speed is the limitation.
Firewire 800	100	72 Read/57 write	800	576 Rd/456 write ***	131 Rd/ 103 Wr	Macworld Practical Figures
Sata 3.0	600	N.A.	4800	N.A.	1092	This is the theoretical interface capability
Sata 3.2	1969	N.A.	15,752	N.A.	3584	This is the theoretical interface capability
Thunderbolt (single hard drive)	1280	112	10,240	896	203	One hard drive – same as USB 3.0. With fast enough drive, Thunderbolt could theoretically deliver 2,330 st. 2496 tracks!
Thunderbolt 2	2560	N.A.	20,480	N.A.	4660	Theoretical interface capability
USB 2 Flashdrive Copies to Itself	N.A.	4	N.A.	32	7	Simultaneous Read/Write. Doubtful this many simultaneous tracks would work in real time
PC Read Internal SATA SSD	N.A.	71	N.A.	568 Read	129 Rd	Windows 7 Pro, Internal SATA SSD. Sandra Light Benchmark. 1 GB file. Intel 7 series Chipset
On Mac, Same file, same SSD Drive via ESATA PCI card	N.A.	21 Rd/5.5 Wr		168 Rd/44 Wr	38 Rd/10 Wr	Same Mac, OSX, XBench
Mac Book Pro SSD Internal Drive	N.A.	412	N.A.	3296	750	XBench. This is incredible. Who needs more? 1GB File. OSX
iMac (1GB File) Promise Raid	N.A.	225 Rd/Wr simult.	N.A.	1800 Rd/ Wr simult.	409	Copy file from RAID to RAID, simultaneous Rd/Wr (stopwatch)
iMac Promise Xbench	N.A.	32 Rd/103 Wr"	N.A.	256 Rd/824 Wr	58 Read/187 Write	Promise Raid 5 is about 2x faster than Internal SSD
iMac Internal SSD OSX	N.A.	17.5 Rd/ 48.7 Wr	N.A.	140 Rd/ 389.6 Wr	31 Rd/88 Wr	Xbench Random 4k blocks Read/Write Uncached
Thunderbolt to SSD/SSD to Thund.	N.A.	32.12/162	N.A.	256/1296	58/294	Promise Raid 5 and Internal SSD copy. SSD is apparently much better at reading than writing. 972MB test file, iMac, OSX.
SSD to SSD iMac OSX	N.A.	243	N.A.	1944	55 simultaneous Rd/Wr	(972MB test file). Simultaneous Read/Write

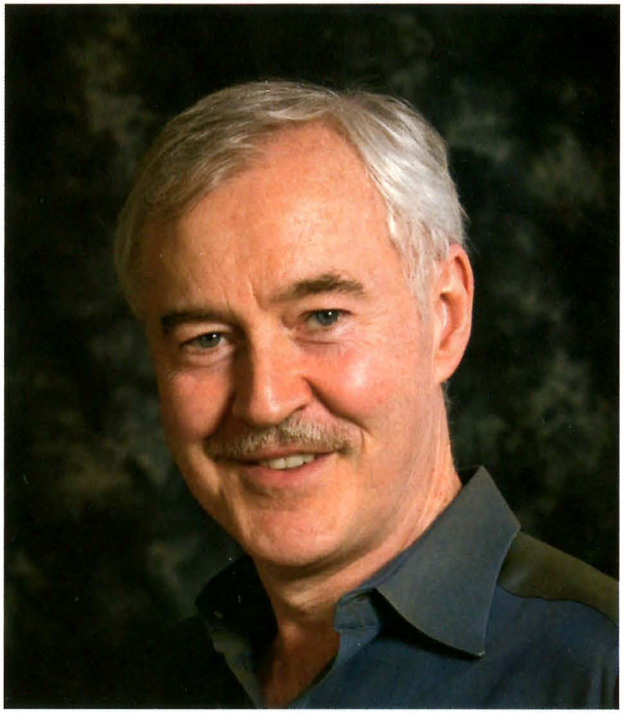
Medium	Theoretical Speed MB/sec	Practical Speed MB/sec	Theoretical Speed Mb/sec	Practical Speed Mb/sec	Practical Stereo 2496 Tracks Real time**	Notes*
SSD to SSD PC (6GB test file)	N.A.	176	N.A.		319 simultaneous Rd/Wr	Practical simultaneous Rd/Wr using Internal SSD
Thunderbolt to Thunderbolt Promise iMac OSX	N.A.	33 Rd/Wr	N.A.	264 Rd/ Wr	60	Practical simultaneous Rd/Wr using Raid 5. 972 MB test file.
USB 3 Seagate External HD and SSD PC (6GB test file)	N.A.	77 copy USB to SSD/170 copy SSD to USB	N.A.	616 copy USB to SSD/1360 copy SSD to USB	140 USB "play", SSD "record"/309 SSD play, USB Record	If the USB is either reading or writing, but not both, it's much faster
USB 3 Seagate External HD to itself PC	N.A.	32 simulta- neous Rd/ Wr	N.A.	256 simul- taneous Rd/Wr	58	Practical simultaneous Rd/Wr (stopwatch). 6GB test file

* GB = Gigabytes, based on 1 TB = 1024 GB. 1 GB = 1024 MB, 1 kB = 1024 B. MB = Megabytes, Mb = Megabits (8 bits/byte). Even the fastest interfaces are limited by the media connected to them. To measure the practical speed of the interface, use the fastest media available. And be aware that media performance may be limited by the interface to which it is connected.

* * Multiply by 2 to get number of mono 2496 tracks. The equivalent number of audio tracks is calculated based on the bitrate we measured while copying a single non-audio data file. It does not take into account the overhead of a DAW to manage actual audio tracks. The more tracks, the more the head has to move around on the hard disk, so the actual is much less than the theoretical. SSDs can handle far more audio tracks than physical hard disks. To be safe, derate these values significantly by a factor of 2-3. Theoretical number of stereo 2496 tracks that the medium could deliver is based on 4,608,000 bits/s, which is 4.39 Mb/s. For serving audio, RAID 5 systems are the safest but not the fastest. Benchmarks were measured with SI Software Sandra Lite, and copy tests measured with a stopwatch. Notice that some media are better at writing (recording). The SSD we tested was much better at reading than writing. In tight situations it pays for you to do a benchmark stopwatch test with a large test file. For some reason we also noticed the Mac in general to be slower than the PC, even with Thunderbolt.

*** When separate read and write figures are given (not performed simultaneously), be aware that simultaneous read and write may be less than that performance. I suggest derating each figure by 3 to be safe.

Appendix VIII. Christopher Morgan Biography



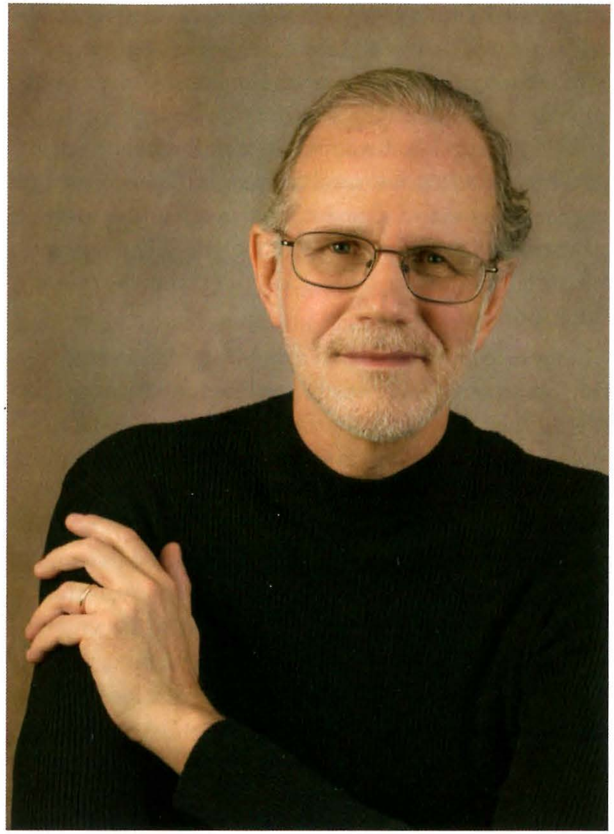
Christopher Morgan is a computer scientist active in the Boston-area hi-tech industry for over three decades. He has degrees in electrical engineering and computer science, and was Editor in Chief of *BYTE* magazine as well as Vice President of Lotus Development Corporation. He is an audiophile and a folk musician, and has been a member of the Audio Engineering Society (AES) and the Institute for Electrical and Electronics Engineers (IEEE). His books include *Wizards and Their Wonders*, about the top 100 people in the hi-tech field, and *The Computer Bowl Trivia Book*.

He is an organizer of the biennial "Gathering for Gardner" conference in Atlanta, where magicians, scientists, mathematicians, and artists gather to share the recreational side of mathematics, and has twice hosted the International Puzzle Party, an annual international convocation of puzzle designers and collectors. He is past president of the Ticknor Society, a Boston-based group of book lovers and collectors, and has served on the boards of the Boston Computer Museum, the Boston Museum of Science, and the Boston Lyric Opera. He is currently working on a book about Lewis Carroll's games and puzzles. Morgan is also a professional magician and photographer *(www.morganpix.com)*.

Appendix IX. Bob Katz Biography

From his earliest years, Bob has been as curious as a Katz. He voraciously reads audio books, service manuals, product spec sheets, license plates, and bumper stickers. But his favorite reads are Science Fiction writers Spider Robinson and Frederick Pohl, which may explain Bob's punny personality. In his teens he dabbled in hypnotism and magic. Bob is an animal lover — all dogs and cats love him back.

Coming from a family of medical doctors, musicians and composers, Bob gravitated to the B flat clarinet at the age of ten; his aunt, a viola teacher, gave Bob his first lesson in solfège and transposition. At the age of 13, he rebuilt his first tape recorder. After wiring the house for sound, he was forced by his parents to remove the microphones he had secreted throughout the house. Clearly destined for a career in audio, by high school he had begun an amateur recording career, plus studying the sciences and linguistics, practicing French and Spanish and looking for female pen pals on three continents. Perhaps out of default he was voted most versatile in his class. Eventually his language skills would reach the point where he can give seminars in any of three languages.



An enthusiastic young man with a passion for good sound, Bob developed a reputation as an audiophile around Hartford, Connecticut town. The local audio stores regularly invited him over, for Bob is never short of opinions. One day he was asked to audition a new pair of speakers with the designer present. After hearing a few notes, Bob ran out of the store covering his ears! Over the years, he has learned to be more diplomatic, but his opinions continue to be defined by a love for the art of audio.

In college he played in an ad hoc Dixieland ensemble, and the treat of his performance life was soloing Sweet Georgia Brown before the homecoming football crowd. Two years at Wesleyan University were followed by two more at the University of Hartford, studying Communication and Theatre, but he spent less time in the classroom and more at the college radio station, where he became recording director. A fan of the Firesign Theatre, Bob used to write and edit humorous radio ads, and he became a DJ, manning a free-form-progressive rock radio show titled *The Katz Meow*, and doing a stint on the commercial rock station.

Bob taught himself analog and digital electronics, and was influenced by a number of creative designers. In Hartford, Bob's mentor was Steve Washburn, an EE who invented a way to nearly double the power-handling of a Hartley 24" woofer and also constructed Bob's first custom-built portable audio console. Bob decided to take a job offer rather than finish his last few college credits, and became (1972) Audio Supervisor of the *Connecticut Public Television Network*, producing every type of program from game shows to documentaries, music and sports. He also learned to mix all kinds of music live. When he wasn't working television, he was on location, recording music groups direct to 2-track.

In 1972, Bob wrote his first article for *dB* magazine, describing a set of mike heaters he developed to warm his AKG microphones and keep them from sputtering due to changes in humidity. This spiked a heated controversy as Stephen Temmer of Gotham Audio wrote a response stating that "Neumann microphones are never affected by humidity" but Bob's experience was supported by some others and in those pre-internet times the controversy remained of modest proportions. Hooked by the writing bug, Bob is a natural-born teacher who puts himself in the mind of the learner. He has been a columnist for *Resolution* Magazine and has written many articles and reviews in publications such as *Byte, dB, RE/P, Mix, AudioMedia, JAES, PAR,* and *Stereophile.*

In 1977 he moved from Connecticut to New York City, and began a recording career in records, radio, TV, and film as well as building and designing recording studios and custom recording equipment. Long before the advent of the home PC, Bob taught himself several computer languages, and sold a 99 page assembly-language program that was used in an embedded system at a brokerage firm. During the primitive time before cell phones, the voice of *Matilda* became well known. Matilda answered Bob's phone and forwarded calls to any place Bob happened to be. Visitors to Bob's house were dismayed to discover that sultry-voiced Matilda was not flesh and blood but rather a 6502-based controller, DTMF encoders, decoders and other gear. Matilda's true identity remains a mystery today.

From 1978-79, he taught at the Institute of Audio Research, supervised the rebuild of their audio console and studios and began a friendship with IAR's founder Al Grundy, mentor and influence. Other New York era influences include Ray Rayburn and acousticians Francis Daniel and Doug Jones. In the 80s, one of his clients was the spoken-word label, *Caedmon Records*, where he recorded actors including Lillian Gish, Ben Kingsley, Lynn Redgrave and Christopher Plummer.

An active member of the *New York Audio Society*, Bob was the ultimate audiophile. This led to a full-page interview/article in the *Village Voice* called *Sex With The Proper Stereo*, a story about Bob's railroad apartment on East 90th with the empty refrigerator in the kitchen and mysterious monoliths in the living room.

But the refrigerator was not empty for long. In 1984, Bob was doing sound for a motion picture in Venezuela and met multi-lingual Mary Kent, production assistant. After the filming, Bob invited Mary to come to New York for a vacation that became a permanent engagement! One day new girlfriend Mary came home and turned on the stereo system in the wrong order, blowing up the Krell amplifier and one of the Symdex woofers producing sparks and blue smoke. When Bob arrived home, he calmed her down — "Don't worry, Mary, your love for me means more than any stereo system." Bob and Mary have been together ever since (Mary jokes that she's really in love with the stereo system).

One day Bob received a call from musician David Chesky, who had read the *Voice* article and was looking for an audiophile recording engineer. In 1988 this led to a long and pleasant association with Chesky Records, which became the premiere audiophile record label. Bob specializes in minimalist miking techniques (no overdubs) for capturing jazz and other music that commonly is multimiked. His recordings are musically balanced, exciting and intimate while retaining dynamics, depth and space. In *1989* he built the world's first working model of the DBX/UltraAnalog 128x oversampling A/D converter, and produced the world's first oversampled commercial recordings. Over the years, the converter was refined, until by *1996* Bob found a commercial model that performed slightly better. Bob has recorded about 150 records for Chesky, including his second Grammy-winner, and in *1997* the world's first commercial 96 kHz/24 bit audio DVD (on DVD-Video).

This obsession with good sound has developed into Bob's passion: *Audio Mastering*. Daily, he applies his specialized techniques to bring the exciting sound qualities of live music to every form recorded today. In *1990* he founded *Digital Domain*, which masters music from pop, rock, and rap to audiophile classical. Besides mastering, Digital Domain provides complete services to independent labels and clients, graphic design and replication. Mary, who became Bob's wife, is an accomplished photographer and graphic artist, the visual half of the Digital Domain team and more than two-thirds of the charm. In *1996*, Bob and Mary moved the company from New York to Orlando, adding numerous Florida-based artists and labels to the international clientele.

In the *90s*, Bob invented three commercial products, found in mastering rooms around the world. The first product, the *FCN-1 Format Converter*, was dubbed by Roger Nichols the "Swiss-Army knife of digital audio". Then came the *VSP model P* and *S* Digital Audio Control Centers, which received a Class A rating in *Stereophile* Magazine. These devices perform jitter reduction, routing, and sample rate conversion.

Bob has delivered lectures and seminars to the Audio Engineering Society at conventions, sections and chaired AES workshops. He was Convention Workshops Chairman, Facilities Chairman and Chair of the AES New York Section. In *1991*, Bob began the *digido.com* website, second audio URL to make the World Wide Web. An education-oriented site, it has become a premium source for audio information. Thousands of pages around the globe have linked to www.digido.com. The desire to pass experiential wisdom to fellow musicians, producers and engineers inspired the production of two books, this tome and *iTunes Music*.

Bob's first 21st century invention has received a U.S. Patent. He designed and introduced an entire new category of audio processor, the *Ambience Recovery Processor*, which uses psychoacoustics to extract and enhance the existing depth, space, and definition of recordings. UAD, Weiss Audio and Z-Systems have licensed Bob's K-Stereo™ and K-Surround™ processes.

Bob has mastered CDs for labels including EMI, BMG, Fania, Virgin, Warner (WEA), Sony Music, Walt Disney, Boa, Arbors, Apple Jazz, Laser's Edge, and Sage Arts. He enjoys the Celtic music of Scotland, Ireland, Spain and North America, Latin and other world-music, Jazz, Folk, Bluegrass, Progressive Rock/Fusion, Classical, Alternative-Rock, and many other forms. Clients include a performance artist and poet from Iceland; several Celtic and rock groups from Spain; the popular music of India; top rock groups from Mexico and New Zealand; progressive rock and fusion artists from North America, France, Switzerland, Sweden and Portugal; Latin-Jazz, Merengue and Salsa from the U.S., Colombia, Cuba, Puerto Rico and Venezuela; Samba/pop from Brazil; tango and pop music from Argentina and Colombia, classical/pop from China, and a Moroccan group called *Mo' Rockin'*.

Bob has mastered several Grammy-winnng recordings, including: *Olga Viva, Viva Olga*, by the charismatic Olga Tañon, which received the *Grammy* for Best Merengue Album, *2000*; *Portraits of Cuba*, by virtuoso Paquito D'Rivera, whch received the *Grammy* for Best Latin Jazz Performance, *1996*; and *The Words of Gandhi*, by Ben Kingsley, with music by Ravi Shankar, which received the *Grammy* for Best spoken word, *1984*. In *2001* and *2002*, the

Parents' Choice Foundation bestowed its highest honor twice on albums Bob mastered, giving the Gold Award to children's CDs *Ants In My Pants*, and *Old Mr. Mackle Hackle*, by inventive artist Gunnar Madsen. The Fox Family's album reached #1 on the Bluegrass charts. African drummer Babatunde Olatunji's *Love Drum Talk*, *1997*, was Grammy-nominated. Erin Bowman's single *Kingboy* was the fifth most-played song on satellite radio in 2013.

Bob's recordings have received *disc of the month* in *Stereophile* and other magazines numerous times. Reviews include: "best audiophile jazz album ever made" (McCoy Tyner: *New York Reunion* reviewed in *Stereophile*). "If you care about recorded sound as I do, you care about the engineers who get sound recorded right. Especially you appreciate a man like Bob Katz who captures jazz as it should be caught." (John Pizzarelli, *My Blue Heaven* reviewed in the San Diego Voice & Viewpoint). "Disc of the month. Performance 10, Sound 10" (David Chesky: *New York Chorinhos*, in *CD Review*). "The best modern-instrument orchestral recording I have heard, and I don't know of many that really come close." (Bob's remastering of Dvorák: Symphony 9, reviewed in *Stereophile*).

Some of the great artists Bob is privileged to have recorded and/or mastered include: Juan de Marcos González and the Afro-Cuban All Stars, Monty Alexander, Carl Allen, Jay Anderson, Lenny Andrade, Michael Andrew, Issa Babayago, Ray Barretto, Lucecita Benitez, Berkshire String Quartet, Ruben Blades, Everton Blender, Gordon Bok, Luis Bonfa, Boys of the Lough, Bill Bruford, Irene Cara, Ron Carter, Brandi Carlile, Cyrus Chestnut, George Coleman, Willie Colon, Larry Coryell, Counting Crows, Celia Cruz, Joe Cuba, Eddie Daniels, Los Dan Den, Dave Dobbyn, Paquito D'Rivera, Arturo Delmoni, Garry Dial, Dr. John, Eastern Rebellion, Toulouse Engelhardt, Robin Eubanks, George Faber, John Faddis, Fania All Stars, David Finck, Tommy Flanagan, Foghat, Fox Family, Johnny Frigo, Kevin Gilbert, Ian Gillan, Dizzy Gillespie, Whoopi Goldberg, Ricky Gonzalez, Bill Goodwin, Arlo Guthrie, Steve Hackett, Lionel Hampton, Larry Harlow, Emmy Lou Harris, Tom Harrell, Hartford Symphony, Jimmy Heath, Levon Helm, Vincent Herring, Conrad Herwig, Jon Hicks, Billy Higgins, Milt Hinton, Fred Hirsch, Freddie Hubbard, Garth Hudson, David Hykes Harmonic Choir, Dick Hyman, Ahmad Jamal, Antonio Carlos Jobim, Clifford Jordan, Sara K., Connie Kay, Ali Khan, Kentucky Colonels, Lee Konitz, Hector Lavoe, Hubert Laws, Peggy Lee, Chuck Loeb, Joe Lovano, La Lupe, Patti Lupone, Gunnar Madsen, Jimmy Madison, Taj Mahal, Sean Malone, Manhattan String Quartet, Herbie Mann, Michael Manring, Marley's Ghost, Winton Marsalis, Dave McKenna, Jackie McLean, Jim McNeely, Milladoiro, Mississippi Charles Bevels, Los Mocosos, Max Morath, Paul Motian, Necrophagist, New England Conservatory Ragtime Ensemble, New York Renaissance Band, Sinead O'Connor, Johnny Pacheco, Eddie Palmieri, Van Dyke Parks, Gene Parsons, Gram Parsons, Danilo Perez, Itzhak Perlman, Billy Peterson, Ricky Peterson, Bucky Pizzarelli, John Pizzarelli, Chris Potter, Tito Puente, Kenny Rankin, Richie Ray and Bobby Cruz, Mike Renzi, Rincon Ramblers, Sam Rivers, Red Rodney, Rodrigo Romani, Michael Rose, Phil Rosenthal, Rey Ruiz, Mongo Santamaria, Horace Silver, Paul Simon, Lew Soloff, George 'Harmonica' Smith, Soneros Del Barrio, Spanish Harlem Orchestra, Janos Starker, Olga Tañon, Ben Taylor, Livingston Taylor, Clark Terry, Thad Jones/Mel Lewis Big Band, Turbulence, Steve Turre, Stanley Turrentine, McCoy Tyner, Jay Ungar, U.S. Coast Guard Band, U.S. Marine Band, Amadito Valdez, Kenny Washington, Peter Washington, Doc Watson and Son, Clarence White, Widespread Jazz Orchestra, Robert Pete Williams, Larry Willis, Cassandra Wilson and Phil Woods.

—by Mary Kent (who knows him best)

Appendix X. Photo Credits and Notes

All photos (except where noted)
© Mary Kent Photography, www.marykentphoto.com. Screen shots used either with permission or according to manufacturer's guidelines; product shots provided by the manufacturers or from manufacturer websites.

Pages 4-5: Hand with iPad and console, Orlando, FL
Page 6: Beach Photo. Captiva Island, Florida
Page 25: Amanda Shelton with headphones, New Smyrna Beach, Florida
Page 37: Bob Katz in front of acoustic amplifier at Parque de los Deseos, Medellin, Colombia
Pages 38-39: Detail (stretched perspective), Parthenon, Athens, Greece
Page 40: Silver Spurs Rodeo, Kissimmee, Florida. Sometimes our mastering day is a wild ride!
Page 55: Sun and Fun, Lakeland, FL. Top: Original image of balloon; Bottom: Enhanced (equalized) image of the same balloon.
Pages 72-73: New York City. Photographic range from light to dark, small to large representing *audio dynamic range.*
Page 80: David Vogel, character actor (squeezed perspective), representing *compression.*
Page 101: U.S. Navy Blue Angels, Sun and Fun, Lakeland, FL. Representing *upward expansion*.
Pages 110-111: Lovely Santorini Island, Greece. Left side represents the "noisy" or "uncorrected" scene.
Page 125: Digital Domain Studio A, north of Orlando, FL. Wide and punchy!
Pages 140-141: Red Rock Canyon, north of Las Vegas, NV, presenting multiple layers of depth.
Pages 152-153: Photos of surround sound engineers provided by the participants.
Chapters 12-13, pages 170-185: Some photos taken in Digital Domain Studio A, most provided by manufacturers or from manufacturer websites.
Pages 186-187: Hotel restaurant, Medellin, Colombia
Pages 188-189: Fisherman's rope, Pontevedra, Spain
Page 197: Bob's custom bitscope, photo by Bob, Digital Domain Studio A, north of Orlando, FL
Page 198: Beautiful Brianna Peterson-Magly, audio engineer, Orlando, FL. Left half represents truncated (undithered) information. We prefer her dithered side!
Pages 214-215: Pleiades Star Cluster, deep space, photo by Hubble Space Telescope, NASA, public domain
Page 215: Pyrex measuring cup, Orlando, FL
Page 217: Faucet, photo by Bob Katz, Orlando, FL
Page 239: Stove top, restaurant, Orlando, FL
Pages 240-241: WWII fighter planes at Sun and Fun, Lakeland FL
Page 257: David Vogel, surprised it sounds so loud, in Orlando, FL
Page 262-263: Bob Katz at music festival near Orlando, FL, measuring SPL with iPhone app *Audio Tool.* Forte measures 84 dB, right on the money!
Page 273: The monitors at Phat Planet Studios, Orlando, FL
Page 279: Alexander Vyverman, head of audio department at SAE Institute, Brussels, shows it's important to protect your ears when taking measurements!
Page 293: A single modern-day audio circuit board has more transistors than were used for the first moon launch. This photograph has been processed with trillions of transistors in Adobe Photoshop.
Page 310: Photo by Bob Katz. A hummingbird in Napa Valley, California, hovers next to a flower to capture its nectar by beating its wings perhaps as fast as 200 beats per second. Despite a shutter speed of 1/1000 second,

the action is not completely frozen. Actually the slightly blurred wings are quite beautiful in contrast to the sharp flower. This photograph represents the need for a high sample rate to capture high frequency information.

Page 321: This double exposure of a stop sign presents the message that jitter should not be ignored!

Page 338: Space Shuttle *Discovery*. We think the date is March 7,2001, one day before a launch from historic launchpad 39B at Kennedy Space Center. Audio is not rocket science, but sometimes it seems that way to audio engineers!

Page 346: Can't fix a broken CD once it's in the case!

Glossary

3 To 1 Rule A recommendation, first defined byLou Borroughs: The distance between microphones should be three times the distance between each microphone and the source of the sound to which it is being applied, to avoid comb filtering. See Chapter 10.

A

A-Weighting See Weighting.

Absolute Polarity Standard *AES 26-2001* states that microphones must produce a positive-going voltage on pin 2 when excited with an *acoustic compression* — an increase of the instantaneous sound pressure that causes displacement of the microphone diaphragm away from the sound source. Loudspeakers should move toward the listener when excited with a positive-going voltage on the loudspeaker terminal which is marked *positive*. See Chapter 9.

Acoustic Compression See *Absolute Polarity*.

ADAT An obsolete digital tape format. The ADAT digital interface remains a standard used today, on an optical connector.

ADC Analog-to-digital convertor, a circuit that converts continuous signals, coming from the analog domain, into discrete digital numbers.

AES-3ID A digital audio interface that employs a BNC connector, with 1 volt peak to peak level. The protocol is identical to AES/EBU.

AES/EBU Digital audio interface jointly conceived by the Audio Engineering Society and the European Broadcasting Union. It is limited to 24 bits/channel and stereo.

AGC Automatic Gain Control. Compression that brings up low-level passages. See Chapter 7.

AIFF (along with WAVE, BWF, SD2, MP3, AAC), a type of audio file format. SDII is obsolete, please do not use.

Album Normalization A method of adjusting the loudness of program material so that the loudest song in the album is adjusted to *target level* and all other songs retain their relative levels as they were in the original album before normalization. See *Target Level*. Also see Chapters 16 and 17.

Aliasing A beat note or difference frequency between the audio content and the sample rate, a form of intermodulation distortion. Proper filtering should eliminate aliases, but see Chapter 22.

Amplitude Level of the audio signal.

Amplitude Masking occurs when a louder sound masks (hides) a softer one, especially if the two sounds lie in the same frequency range.

Anti-Image Filter A filter in a DAC to prevent "images" of the baseband signal from appearing at frequencies far above the baseband (See Chapter 23).

Archive A backup on a medium that is supposed to last a long time (30 years or more).

ASRC Asynchronous sample rate converter. A converter from one sample rate to another which can work with a wide relationship of input to output frequency, and thus can deal with varispeeded rates. Filter coefficients are continuously variable, computed on the fly. However, this may not yield the lowest distortion. Chipset-based ASRCs may not have the best internal resolution. See Chapter 24.

Asynchronous (Communication) Non-clocked connections between devices. See Chapter 24.

Asynchronous Router A digital audio router which does not require clocking. It can contain signals which are at different sample rates and can switch virtually any type of signal. See Chapter 14.

ATSC Advanced Television Systems Committee. A U.S.-based organization that sets standards, including the new loudness measurement and normalization standards.

Attack Time The time it takes for the compressor to implement full gain reduction after the signal has exceeded the threshold. See Chapter 6.

Attenuation When expressed in dB is an optional term for negative gain, e.g., a loss. Example: 20 dB attenuation is the same as -20 dB gain.

Audible Resolution The lowest signal which can be heard above the noise.

Authoring The process of recording source material (audio, video or other data) onto the release format, which may include adding menus, and interactive content. To author, one has to know all the rules and specifications of the format, e.g. DVD, Blu-Ray.

AVG Average (abbrev.).

Azimuth The timing relationship between channels in an analog tape deck, affected by the angle of the head.

B

Backup To place copies of files in a safe place. Since you should "never turn your back on computers," says Bob Ludwig, and "you're only one mouse click away from disaster," says Bob Katz, then "you're never backed up until your data is in at least three places," says George Massenburg!

Balanced Interface An interface with each polarity of the signal separated from ground and at opposite polarities from each other. For example, in a balanced XLR, when the signal is positive at pin 2, it is negative at pin 3 and vice versa.

Bass Management A crossover which directs bass information to the subwoofer. See Chapter 21.

Baxandall (after Peter Baxandall). A gentle equalization curve with its highest or lowest level at the frequency extremes. See Chapter 4.

Bell Curve A parametric EQ curve shaped like a bell. See Chapter 4.

Bit Depth See *Wordlength*.

Bit-Transparent The output is a perfect digital clone of the input, including the source wordlength.

Bitrate Speed in number of bits per second. As distinguished from word-length. In linear PCM, bitrate is the product of the wordlength (sample size) and the sample rate.

BLER Block Error rate. A count of the number of errors on an optical disc. See Chapter 1.

Blu-Ray An optical disc standard that can hold up to 25 or even 50 GB of information in multiple layers. This permits encoding very high resolution video and audio with no compromise.

Bonger (also known as *Gonger*). A test signal invented by Chris Travers of the BBC. Excellent for revealing distortion anywhere in a signal chain.

Bounce To copy the output of a workstation to a new file.

Buzz Power-line related noise predominantly at the higher harmonics of the power line frequency. See Chapter 8.

C

C1/C2 Errors Soft errors on CD-DA are correctable in two layers of defense: C1 and C2. Hard errors are known as CU. If the C1 correction fails, C2 takes over, and if that fails, a CU error occurs and the player goes into error concealment.

CALM Act Commercial Advertisement Loudness Mitigation Act. A U.S. regulation for the loudness of television commercials. But since this implies that the program level is already at some kind of standard, it led to the ATSC loudness standard. See Chapter 16.

CD-DA See "Compact Disc Digital Audio."

Clipping (Noun: **Clipper**). A form of distortion that happens when an amplifier is asked to create a signal greater than its maximum capacity. When trying to go above the maximum capacity of the amplifier, the signal is said to be "clipped". In digital, the maximum capacity is known to be be 0 dBFS, and any overs will cause distortion to appear on its outputs.

Codec (Coder-Decoder). An algorithm that performs encoding (recording) and decoding (playback) on a digital data stream or signal. There are both lossy and lossless codecs. WAV, AIFF are examples of lossless codecs while mp3 and AAC are lossy. The sound quality of a lossy codec is dependent on the algorithm and bitrate.

Coincident Microphones A simple miking technique which places the heads of each microphone very close to one another. See Chapter 10.

Comb Filtering A frequency response anomaly introduced when combining an audio signal and a delayed replica to the same output channel. See Chapter 2.

Compact Disc Digital Audio (CD-DA) A 16-bit stereo 5" disc standard jointly developed by Sony and Philips in 1980. This is also known as the *Red Book standard*. See also *Red Book* for other forms of the compact disc such as the CD-ROM, which carries files.

Companion Compression followed by complementary expansion. See Chapter 7.

Complementary noise reduction systems work in two parts, a compressor and a separate expander. E.G. The Dolby System.

Compression Ratio The ratio between input and output level of a compressor above the threshold point. See Chapter 6.

Compression Reduction of dynamic range. See Chapter 5.

Convolution A mathematical way of combining two signals to form a third signal.

Crescendo (plural: **Crescendi**). A gradual increase in loudness over time.

Crest Factor The difference between the peak amplitude of a waveform and its RMS value. For actual musical signals, use the new term *PLR* instead.

D

DAC Digital-to-analog convertor, a circuit that converts discrete digital numbers, into continuous signals (a voltage) in the analog domain.

DAT Digital Audio Tape Recorder. Obsolete.

DAW Digital Audio Workstation. Usually a computer with dedicated hardware and software for editing and processing digital audio.

dB Decibels. A logarithmic measure of audio level. See Chapter 16.

dBFS Decibels relative to full scale. 0 dBFS means "0 dB reference full scale," as on a digital meter. For example, -10 dBFS is a level 10 dB below full scale. See Chapter 16.

dBm Decibels relative to one milliwatt.

dB SPL Decibels relative to 0 dB SPL (Sound Pressure Level).

dBTP Decibels relative to true peak level.

dBu Decibels relative to 0.775 volts, means "Decibels unterminated"

DDP Disc Description Protocol image file, sometimes abbreviated DDPi, which can be placed on data disc or uploaded to the plant via FTP. Used for sending CD-DA masters to replication plants. Also used for DVD.

Decrescendo (plural: **Decrescendi**). A gradual decrease in loudness over time.

Decimator A digital filter that reduces sample rate, usually incorporated inside ADCs.

Declicker A noise reduction process dedicated to removing short duration clicks. Sometimes crackle can be removed with a declicker, and sometimes clicks can be removede with a decrackler. See Chapter 8.

Declipper Clipping that is level-dependent and has a clearly-defined threshold is the easiest to repair. Dedicated declippers, such as Cedar's Declip and Izotope's Declipper, remove (or at least reduce) clipping distortion by interpolating the missing pieces. See Chapter 8.

Decrackler A noise reduction process dedicated to removing clicks which are so numerous they are known as *crackle*. It can also make an excellent distortion-softener or remover, when selectively applied. See Chapter 8.

De-Esser A device dedicated to reducing excessive sibilance (S sounds). Can also be used to sweeten or soften harshness or bright instruments as a form of dynamic equalization.

Delay Mixing Adding a delay to each close mike to synchronize it with the main pair, helps to pull the soloist back and helps to maintain natural depth. See Chapter 10.

Dereverberator A process designed to reduce the reverberation in a recording which presumably has too much. Be sure to listen for artifacts which, depending on the algorithm which is used, could produce noises much like a codec's "space monkeys."

Descratcher A noise reduction process dedicated to removing scratches, usually found on LP records. Also note that a descratcher can make an excellent distortion-softener or remover, when selectively applied.

Dialnorm Dolby's normalization standard, based on dialogue level. -31 dBFS is the lowest dialnorm level, which means 0 dB attenuation would be applied during normalization. Therefore a program with a dialnorm level of -21 dBFS would be attenuated by 10 dB during normalization.

Digititis My term: the inharmonic distortion caused by artifacts of digital audio processing, perceived as a harshness in the sound. Technically it is inharmonic distortion, the beating of naturally-occurring harmonics against the sample rate. This kind of distortion can occur in any digital compressor or limiter. Clipping can also cause digititis. Oversampling helps reduce digititis (See Chapter 22).

Directional Masking The hiding of one sound by another which occurs in the same physical location. For example, reverberation coming from the same direction as the direct sound can be masked. See Chapter 10.

Disc at Once A continuous, non-stop mode of CD writing suitable for creating a master disc.

Distortion Although a layperson would lump distortion and noise together, an audio engineer characterizes distortion as a particular form of noise: one that is correlated with the signal. Distortion can be low level and sound much like what is normally called noise, or it can be high level and quite obtrusive, lying on the peaks of the signal.

Dither A process that linearizes digital audio by adding a random noise signal at the point just before wordlength truncation. Dither is required for clean digital audio recording and processing. After dithering, the wordlength can be safely truncated or shortened, but truncation without dithering results in quantization distortion. See Chapter 15.

Downsampling Reducing the sample rate.

Downward Compression The most popular form of dynamic modification. It brings high-level passages down. Limiting is a special case: it is downward compression with a very high ratio. See Chapters 5-7.

Downward Expansion A process which lowers the level of low passages. Most downward expanders are processors employed to reduce noise, hiss, or leakage. See Chapters 5-7.

DSD (Direct Stream Digital). A digital format that uses delta-sigma modulation, also known as one-bit encoding. It is the audio format used on the SACD (Super Audio Compact Disc), as opposed to multibit PCM, the most common digital audio format.

DSP (Digital Signal Processing). The processing of a stream of information by performing numerical calculations.

Duplication Reproducing optical discs by copying multiple writable discs from a master CDR or DVD-R to multiple writers. In contrast to *replication*.

DVD-A DVD originally stood for Digital Video Disc, but it has been redubbed *Digital Versatile Disc* since it can support computer, audio, and video formats. The -A suffix defines the multichannel audio disc standard that supports a wide range of PCM sample rates and wordlengths, and limited graphics. DVD-As are functionally obsolete since the Blu-Ray is much more versatile.

DVD-V A video and audio disc standard that supports multichannel digital audio sample rates up to 48 kHz/24-bit, and 2-channel digital audio at 96 kHz and 192 kHz, but there is usually not enough room on the disc to fit high-quality video and high resolution audio at the same time. Functionally replaced by Blu-Ray.

Dynamics Processing The artistic and technical tasks of leveling the audio.

Dynamic Range The range in decibels between the highest and lowest recorded levels. This has been formally defined by the ITU and EBU with *LRA, loudness range*. LRA is now the preferred term because it is a standardized measurement.

E

E-E Pronounced "E to E." Electronics to electronics. For example, when a recorder or DAW is put into record, its output can monitor its input directly. Also known as *input mode* or *input monitor*.

EAN Code (Mode 2 data) A barcode that contains information about the product, usually the entire record album, containing 13 digits. Supersedes the old UPC code, which was only 12 digits. See Chapter 1, page 16.

EBU European Broadcasting Union.

EDL Edit decision list. Also known as Playlist. Instead of cutting the actual audio, an EDL is a list of instructions where and how to cut and reproduce the audio when played back. Thus, many different versions or playbacks of the same audio can be reproduced from the audio files. An EDL is to audio as a Word Processor is to words.

EFM (Eight-to-Fourteen Modulation), used for channel encoding on CDs. EFM breaks the data (PCM audio in case of the CD, together with data bits) into 8-bit blocks (bytes). Each 8-bit block is translated into a corresponding 14-bit word using a lookup table.

Emphasis In an effort to improve the already excellent signal-to-noise ratio of the Compact disc, CDs (as well as digital tapes) can be recorded with emphasis. If it is decided to use emphasis, the recording is made with a calibrated high frequency boost (called Emphasis), and during playback, a corresponding high frequency rolloff (called Deemphasis) is applied. Thus, in theory, signal-to-noise ratio is improved, though in practice the loss of high frequency headroom may reduce any audible improvement. Most CDs made today do not use emphasis.

Emulation Digital processor which reproduces another processor (usually an analog model) by using the mathematical transfer characteristics of the source processor, e.g. its ratios, frequency response, distortion characteristics and output levels for each incoming level. Different from convolution, which samples the source processor and convolves that sampled characteristic with the incoming audio. Do not leap to conclusions whether emulation or convolution sounds better, it depends on the skill of the programmer and the device being emulated.

Equal Loudness Contours A measure of sound pressure over the frequency spectrum for which a listener perceives a constant loudness. This was first measured by Fletcher and Munson in 1933, and later researchers' work that became the ISO 226 standard.

Equalization Adjusting the frequency response characteristics of a recording, e.g. making it brighter or duller, or bassier or thinner. See Chapter 4.

ETC Energy-Time-Curve (See Chapter 21).

Expansion An increase of dynamic range.

External Sync A signal which can be applied to a converter or a DAW to lock it to a master clock.

Eye Pattern An oscilloscope or analyzer display which reveals the integrity of a digital interface and how stably it is locked to the source signal.

F

FET Field effect transistor. A device which can be used as a gain reduction element in a compressor/limiter.

Fingerprint Also known as a noise profile. A sample of noise without signal to facilitate noise reduction. See Chapter 8.

FIR Finite impulse response. A type of digital filter. See Chapter 4.

Firewire A high-speed bi-directional serial interface originally developed by Apple computer, but then officially adopted as standard IEEE 1394, for use with digital audio, video, hard drives, controllers, etc.

Fixed-Point Vs. Floating Point Fixed-point notation uses a finite binary number whose range in 16-bit is 96 dB, in 24-bit is 144 dB. Floating point notation can represent thousands of dB of dynamic range through use of exponents. See Chapters 15-16.

Fletcher-Munson See *Equal Loudness Contours*.

Frames There are two commonly used frame standards in CD work, with different lengths: 75 CD Frames in a second, as opposed to 30 SMPTE frames per second. Modern PQ lists are usually expressed in CD Frames, but the older 1630 systems used SMPTE frames, which have less timing resolution.

G

Gain Makeup A simple gain amplifier after compression, since downward compression reduces the level of the loudest passages, gain makeup can be used to raise the average level. See Chapter 6.

Gain Reduction (GR) The meter in a compressor which tells how much the compressor is reducing the gain of the signal. See Chapter 6.

Gain The difference between input and output level. See Chapter 16.

Gating The EBU standard R128 defines a measurement gate that purposely ignores soft program material more than 10 LU below the average, in order to prevent extra-soft passages and fadeouts from over-influencing the true average loudness measure. Most current EBU-compliant meters incorporate the gate. I recommend that all loudness meters for music measurement incorporate the gate.

Glass Master Glass Mastering is the process of transferring the CD master to a physical image of the pits on a coated glass substrate. See Chapter 1.

H

Haas Effect Also known as the Precedence or Fusion Effect. A psychoacoustic effect whereby the ear fuses a delayed signal with its source under very specific conditions of delay time and level. When the delayed signal is too loud or the delay is too long, the fusion breaks down and the ear hears two sources instead of one. See Chapter 10.

Hard Error A media error on a CD which is not technically correctable, but if the error length is short enough it will be interpolated by the player using surrounding audio material. The vast majority of times interpolated hard errors are inaudible to the listener.

Hard Knee See "Knee".

Headroom The distance between program level and the peak capability of the medium. See Chapter 16, page 223 for a diagram.

Hot CD, or Hot Master A recording with a very high program loudness.

Hum The lower frequency components of the power line, usually the fundamental, second and third harmonics. Usually either 60, 120, or 180 Hz or 50, 100 or 150 Hz.

Hypercompression Compression applied for the sake of increasing intrinsic loudness but without caring about the decrease in sound quality or increase in distortion. See Chapter 6.

I

IIR A type of digital filter. See Chapter 4.

Index 0 A "pause" mark in the CD. This is optional, and indicates a space between the end of the previous track and the beginning of the next. Audio is allowed in the pause. See Chapter 3 for examples.

Index 1 The flag which is the start of a track mark in the CD.

Integrated Loudness Also known as *Program Loudness* or *PL*. Loudness as defined by ITU BS.1770-3.

Intensity A measure of energy flow per unit area. For practical purposes, sound intensity is the same as SPL

Internal Sync A crystal clock located inside the converter directly drives the circuitry.

International Standard Recording Code (ISRC) The International Standard Recording Code, provided to record labels by the RIAA, is a unique code for each track on the album. Theoretically this allows automated logging systems to be used at radio stations to track copyright ownership/royalties, but this was only true in the days when radio broadcast music CDs (most radio stations have converted to playing back audio files). Standards bodies are working on a way of placing ISRC in the metadata of broadcast wav (BWF) files. See Chapter 1, page 17.

Isochronous Communication between a converter and the digital audio source where the converter can run on internal sync or wordclock sync. The converter requests samples from the source when it needs them as opposed to being driven by the clock of a source. The term is more refined than this but for purposes of this book the definition is sufficient.

ISRC See "International Standard Recording Code".

Itu Bs.1770-3 by the International Telecommunication Union. The international audio standard for measuring loudness. The -3 refers to the version number.

J

Jitter Timing variations in the digital audio clock, producing distortions. *Interface Jitter*, the jitter in the interconnections between devices, may or may not manifest into *Sampling Jitter*, which is audible. See Chapter 24.

K

K-Stereo™, K-Surround™ Patented processes for extracting and enhancing the already existing ambience of recordings.

K-System An integrated system of metering and monitoring proposed by the author whereby the zero point on the meter is the nominal level and corresponds with a specific SPL when used with a calibrated monitor control. The metering portion of the new loudness-based K-System is a BS.1770-3 loudness meter with an adjustable 0 LU point. The calibrated monitor control is constructed according to the formula that the *sum of the momentary loudness and the monitor control position should be -20 LUFS* (our desired forte). See Chapter 19, pages 265-266.

kHz Abbreviation for kiloHertz, meaning audio frequency in thousands of cycles per second. Commonly this usage also applies to sample rate, which can cause confusion. When there might be confusion, we add the term sample rate where appropriate.

Knee The portion of the curve in a compressor or dynamics processor near the threshold. It marks the transition between unity gain and compressed output. **Soft Knee** is a knee with a gradual transition, and **Hard Knee** is a knee with a sharp, distinct transition. See Chapter 6.

L

Latency Compensation Within DAWs, all outputs of every bus are sample aligned, no matter how many plugins are inserted on that bus.

Least Significant Bit See "LSB".

LEDR Listening Environment Diagnostic Recording, invented by acoustician Doug Jones. A powerful but simple test for playback system and room acoustics accuracy, available on test CD JD37 from Chesky Records. If your system cannot pass the LEDR test, then replace loudspeakers, relocate them and/or work on room acoustics.

Level A measure of intensity, but when used alone it could mean almost anything, so it should be accompanied by a qualifier such as voltage or power. See Chapter 16.

Leveling Adjusting the levels of each song in an album so that they work together esthetically. It does not mean to set the songs to equal measured loudness, or the ballads would sound too loud, for example.

Limiter, Limiting The definition of limiting is really a matter of degree, but most authorities call a compressor with a ratio of 10:1 or greater a limiter.

Look-Ahead (also known as **Preview**). A means of finding and controlling the levels of peaks before they occur. Since it's impossible to look into the future, this is accomplished by measuring in real time, but delaying the audio signal slightly to coordinate with the limiter or compressor's action. This function in a dynamics processor allows very fast, or even instantaneous (zero) attack time, which is especially useful in a peak limiter to prevent overloads.

Loudness The listener's perception of intensity. In 2012 the ITU formalized a loudness measurement standard known as BS.1770-3. See Chapter 16.

Loudness Normalization Gain adjustment of loudness of each song or group of songs, or more generally, an entire radio or television program, to a standardized loudness.

Loudness Range The formal term for dynamic range defined by EBU's R-128 recommendation. Abbreviated **LRA**

Low-Pass Filter An equalization filter that removes high frequencies. See Chapter 4 for its esthetic application in audio equalizers, and Chapter 23 for its technical application in converters.

LRA Loudness Range. A formal definition of dynamic range by the European Broadcasting Union's recommendation R-128

LSB Least significant bit, the bit with the smallest (lowest) analog value in the PCM system.

LU Loudness units. 1 LU change is equivalent to one decibel change.

LUFS Loudness units relative to full scale digital. E.G. 0 dBFS is full scale. For example, -20 LUFS is 20 loudness units below full scale.

M

Macrodynamics My term for changes in loudness perceived on a longterm or average basis, anything longer than perhaps 100-200 ms.

MADI (Multichannel Audio Digital Interface) A communication protocol commonly used in digital audio, that provides serial data transmission over coaxial or fibre-optic cables and supports up to 64 channels, with sample rates up to 96 kHz and bit depths of up to 24 bits per channel.

Manual Compression Moving the fader up or down, or manipulating gain in a workstation.

Manual Limiting Making an edit in a DAW around a short transient and reducing its level, for a brief time, hopefully inconsequential to the listener.

Master The formal term for the final medium which is technically ready for replication or duplication.

Measurable Resolution The lowest signal which can be detected above the noise.

MFIT (Mastered for itunes) Apple's program for producing higher quality coded AAC audio. In general, it means that 24-bit/96 kHz or 24-bit 44.1 kHz audio should be supplied as the premaster.

Metadata Data about data. It is used to facilitate the understanding use and management of data. MP3 uses ID3 metadata, which stores the title, artist, album, etc., with the audio itself. Streaming audio takes advantage of metadata to tell the receiver how much gain to use when normalizing, as well as the title, artist name, genre, etc.

Microdynamics My term for the music's rhythmic expression, transient quality, integrity or bounce, which involves the music's short-term peaks. Probably confined to information shorter than 100-200 ms.

Microsecond (µs) One millionth of a second.

MIDI Musical Instrument Digital Interface.

Momentary Loudness Abbreviated "M". A formal standard for a loudness meter or measure defined by the ITU/EBU, the loudness you hear now. M is averaged over a 400 ms period, which corresponds well with the VU meters many of us are used to.

Monitor Control Position The position of a monitor control marked in decibels, given in dB relative to the 0 dB position (usually marked "0" at the highest position of the control.

Moore's Law The empirical observation made in 1965 that the number of transistors on an integrated circuit for minimum component cost doubles every 24 months, some say as fast as 18 months. It has been extrapolated to information and speed of computers as well.

MP3 MPEG-1 Layer 3. Popular audio encoding format that uses a lossy compression algorithm.

MS Compression Compression of stereo material by separating the M (center) information from the S (side) information. See Chapter 9.

Multiple Miking Use of more than a few microphones to capture an ensemble, usually referred to placing a microphone or several microphones on every instrument. See Chapter 10.

N

Noise Unintended sound that interferes with the perception of the signal. See Chapter 8.

Noise Gate A device which attempts to reduce noise by use of a threshold. Signals which are below the threshold are attenuated. See Chapter 8.

Noise Reduction Removal or reduction of noise. See Chapter 8.

Nominal Level The average or RMS level at which an audio device is designed to operate. As opposed to peak level, which is the highest short term level that the device is capable of producing without distortion. See Chapter 16.

Normalization A automatic adjustment of level. Usually described on a song-by-song basis or album basis. Peak normalization sets all the peaks to the same level and is undesirable for many reasons, as described in Chapter 17. Loudness normalization sets the measured loudness of all the songs to the same level, as described in Chapters 16-18.

Null Test A test for bit integrity and to confirm that two files are identical. Line up the two files in a DAW to the sample and invert one. Mix the two together. If the files are identical, there should be no audible or measurable output.

Nyquist Dr. Nyquist, while working for Bell Labs, discovered the sampling theorem, where he states that a sampled waveform contains ALL the information without any distortion, when the sampling rate exceeds twice the highest frequency contained by the sampled waveform. This theory forms the basis for all PCM-based systems. There can be no information in a sampled system above the Nyquist Frequency, which is 1/2 of the sample rate. Other scientists, notably Shannon, deserve credit for refining the theorem.

O

Optical (abbreviated Opto) A gentle type of compressor detector with very low distortion. Not very speedy, however.

Oversampling Raising the sample rate of material so as to avoid artifacts when digitally processing.

P

Pandora's Box In Greek mythology, Pandora was the woman who opened a box releasing all the evils of mankind, leaving only hope inside once she closed it. I use this to describe what happened when peak normalization practice met up with loudness envy!

Parallel Compression Compression which is mixed in with the unmodified source signal. See Chapter 7.

Parametric An equalization technique invented by George Massenburg in 1967 which provides independent and non-interactive control over center frequency, bandwidth, and level of boost or cut.

Passband The passband is the part of the frequency response which is not filtered or attenuated. See Chapter 23.

Passband Ripple Minute variation in frequency response in the audible band. See Chapter 23.

PDR Program-dependent release. See Chapter 6.

Peaking See "Bell Curve"

Peak Level The highest short term level that a device is capable of producing without distortion. As opposed to average or RMS level, the nominal level at which an audio device is designed to operate. See Chapter 16.

Peak To Loudness Ratio (PLR) The ratio between the highest true peak not exceeding 0 dBTP, and the long-term average loudness of the song or album in LUFS. Proposed by Thomas Lund and not yet standardized.

Phantom Center Phantom image. A virtual image between two sources, e.g. between left and right front loudspeakers.

Picosecond (ps). One millionth of one millionth of one second, or 10-12 second.

Plug-In An extra process which can be inserted into a DAW. Some plug-ins utilize the power of an external DSP card, while others, called native plug-ins, utilize the computer's CPU.

PL Abbreviation for Program Loudness

PLR See *Peak to Loudness Ratio*.

Polarity The quality of having two oppositely charged poles, one positive and one negative. Changing polarity means reversing the positive charged pole with the negative, and vice versa. Analog sound travels through electronics as AC current, where polarity inversion can be applied by switching the positive and the negative wire. See Chapter 9.

Popper A combination audio signal generator and receiver that puts out a polarity test signal, which sounds like a pop. Connect its line out to the feed to the loudspeakers, then its built-in microphone or its line input will detect if the signal path is in correct absolute polarity.

PPM Peak Program meter. The quasi-peak meter, with response time between 6-10 ms, is considered obsolete. The digital sample peak meter is also considered to be an unacceptable measure because true peaks can occur in real world devices that exceed this measure. Thus the true peak measure is preferred.

P Flag A flag in the CD subcode that indicates the start of a track

PQ Coding The Compact disc contains a number of subcode areas, each area is named with a letter, from P to W, with information on track number, timing, and so on. PQ coding is the process of defining where track marks should occur in a CD.

Precision The internal data wordlength within the algorithm.

Preview See "Look ahead".

Program Loudness officially defined by the ITU in BS.1770-3, endorsed by the ATSC and EBU. It is often abbreviated PL—the program loudness of a program over time, expressed in loudness units relative to full scale.

Pultec A brand name of the company Pulse Techniques, formed in the early 1950s and no longer in existence. The much-revered Pultec equalizers have been revived in a few pieces of hardware, and in some plugins that emulate the distortion characteristics of its analog electronics and reproduce its unique curves.

Punch The application of judiciously timed compression or expansion to enhance the power and impact of a production. If compression is not applied well or over-applied, punch can be reduced. See Chapters 6 and 9. See also *Snap*.

PWM Pulse Width Modulation. The chameleon of compressor types. A versatile style of compressor design used in analog compressors, which can create virtually any style of time constant, so the unit can sound like optical, FET, or VCA if desired.

Q

Q The parameter Q is defined mathematically as the result of dividing the center frequency by the bandwidth in Hertz at the 3 dB down (up) points measured from the peak (dip) of the curve.

Q Subcode The Q subcode in the CD contains information such as timing and program length, copy prohibit or permit, emphasis condition, and ISRC codes.

Quality Control (QC) Form of checking and ensuring the product that is being delivered is error-free.

Quantization In an analog system, the signal is continuous1, but in a PCM digital system, the amplitude of the output signal is limited to one of a set of fixed values or numbers. This process is called quantization. Each coded value is a discrete step.

R

RAID Redundant array of independent disks. A system to increase the reliability of a hard disk by duplicating the data across more than one disk. In case one disk fails, the data is still intact. For high capacity disk systems I recommend using RAID 6, which allows for failure of up to two disk drives without losing data.

R-128 A recommended loudness normalization standard set by the EBU. Target level is specified as -23 LUFS. Measurement should be gated as specified in ITU BS.1770. Maximum true peak level should not exceed -1 dBTP.

Recovery Time Synonym for "Release Time".

Red Book A Sony/Philips document which defines the standards for the audio CD. The Blue Book defines enhanced CDs with audio and ROM material. Yellow Book defines CD-ROMs. Green Book defines compact disc

Interactive. White Book defines the Video CD. Orange Book defines CD-R or Recordable CDs.

Redithering When processing digitally and reducing the wordlength, we need to add dither in the digital domain. This is commonly called redithering to distinguish it from the intial dither which may have been required during analog to digital conversion. Do not fall for the misconception that additional dither is unnecessary after initial encoding. See Chapter 15.

Release Time The time required for the signal level of a dynamics processor to return to unity gain after it has dropped below the threshold.

Replication Synonym for pressing; the result is a durable molded metalized circle of plastic, sealed under a coat of protective lacquer, which can last for 100 years or more.

Resolution is an overused term that we must define to make it effective. We define the term resolution to indicate whether a source signal of a given level will be represented in the output. This can be expressed as a number of equivalent bits.

Retouch The name of a trademarked process by Cedar which is used to reducing or removing noise. See Chapter 8.

Ringing A high frequency oscillation, usually caused by filters with very narrow bandwidth or sharp cutoff.

RMS Root-Mean-Square. A method of averaging levels which computes the equivalent power of the material. For all naturally-occurring music, an RMS-responding meter will read several dB below the actual peak level of the music at any moment in time.

S

SACD See *DSD*.

Sample Peak Level The peak value of the digital sample, measured by traditional digital meters. Use true peak level instead.

Schroeder Frequency Point Small-room acoustics is divided into low frequency problems and high frequency problems at the so-called Schroeder Frequency point, below which the room behaves largely modally (room modes are standing waves at particular wavelengths that are integer-related to the distance between walls).

Segue A crossfade between two different types of music, pronounced segway, from the Italian seguire meaning to follow.

Sequencing Putting an album together and spacing it, not to be confused with MIDI sequencing. See Chapters 1 and 3.

Sforzando From the Italian: to play a sudden, strong musical attack.

Shelving An equalization curve that affects the level of the entire low frequency or high frequency range below or above a specified frequency. See Chapter 4.

Short-Term Loudness Abbreviated S, has a time window of 3 seconds in a BS.1770 meter.

Shred A colloquial term used by audio engineers to describe clipping a PCM signal to add distortion and loudness character, then turning down the level to hide the fact that clipping has occurred. Over levels do not show but the distortion remains. This practice is not recommended! The average loudness may go up, but the impact goes down. See Chapter 16.

Sidechain The control path for a dynamics processor. See Chapter 6.

Single-Ended noise reduction systems attempt to separate noise from signal without having a specially-recorded fingerprint or noise sample.

Singles Normalization Also known as *track normalization*. A normalizer that adjusts each song to the same loudness. See Chapter 16.

Snap My term for very short upward-moving dynamic contrast that is so short-term that it is not perceived directly as a loudness increase, although it contributes to the partial loudness and the liveliness of the sound. Snap is short term impact, typically below 100-200 ms. It is the important companion to punch. A good engineer needs to concentrate on both punch and snap. Lose too much of either attribute and the recording suffers. For example, we could call the sound of the beater of the bass drum its snap, and the resonance of its diaphragm and body its punch. Snappy reflects the presence of transients and microdynamics in the sound. See Chapters 9 and 17.

SNR Signal-to-Noise Ratio. SNR can be measured in many different ways. Be sure to ask how it was measured if the weighting and the method are not specified.

Soffit An enclosure for loudspeakers built into a wall. Usually the loudspeakers are flush-mounted to eliminate edge diffraction and the wall is specially treated to deal with lower midrange buildup. The walls are specially-constructed and angled for optimum imaging and the loudspeaker cabinets are isolated to prevent vibration. The protocols and isolation techniques for soffit mounting require extreme expertise to do it right, which means that construction, alignment and equalization should be custom-performed by experienced and trained acousticians and architects.

Soft Error A media error on a CD which is fully correctable with the built-in redundancy. It has no consequence to the listener.

Soft Knee See "Knee".

Sound Check Apple's name for its loudness normalization technology.

Space Monkeys A colloquial term for artifacts of overaggressive denoising or low bitrate codecs.

S/PDIF Sony/Philips Digital Interface. Standard IEC-958 and IEC-60958 defines this interface, usually found on an RCA (coaxial) connector.

Spectragram A display of audio showing three "dimensions" at once, time, amplitude and frequency. High amplitudes are indicated in red, and descending levels in orange, yellow, green, then blue. Read it like a musical score. See page 62.

Specular (adj.) Describes the nature of sound reflections which are mirror-like, sharp and distinct, as opposed to diffuse.

SPL Meter A meter that measures sound pressure level.

SRC (Also abbreviated **SFC**) Sample Rate Converter, or Sample Frequency Converter. See Chapter 23. A Synchronous SRC uses fixed filter coefficients, can only convert between certain fixed rates, e.g. 44.1, 48, 88.2 and 96 kHz, and cannot accept varispeeded sources. See also *ASRC*.

State Machine Any type of processor which produces identical output for the same input data, and which does not look at data timing or speed, but only at the state or recent history of the data. Most digital processors are state machines and thus are completely immune to jitter. See Chapter 24.

Stems Individual components in a mix, which may be used in mastering when there are problems in the full mix. Usually stems are wet. The sum of the stems must equal the full mix. See Chapter 9.

Stop Band The area where a filter cuts off the sound. See Chapter 23.

Subtractive Mixing A mixing technique where we drop the level of a fader to take an element's level down, instead of raising the fader controlling the sound we wanted to raise. An excellent technique to learn because it helps to avoid loudness and level escalation

Synchronous Router A digital routing system which requires clocking and that all signals be at the same sample rate and framed to the identical clock.

T

Target In *loudness normalization*, the loudness level which a normalizer will produce at its output. For example, if the target level is -23 LUFS and the sources program loudness is -13 LUFS, the normalizer will attenuate the input signal by 10 dB.

In *digital room correction*, the frequency response and level to which the loudspeakers are being standardized. For example, a curve that is flat to 1 kHz and rolls off to -7 dB at 20 kHz may be the *target frequency response*.

Temporal Masking A psychoacoustic effect. When two signals are close together in time, one of them may hide (mask) the other. See Chapter 4.

THD Total Harmonic Distortion, a measurement of the harmonic distortion present, used to test equipment performance.

Threshold of a compressor is the level at which gain reduction begins,

Time Correction A DSP process which changes the speed or timing of material without altering its pitch.

Tracking Filter When a tonal noise is varying in frequency, as in the case of analog tapes with varying speed, a special kind of tracking filter is required, usually found in forensic suites.

Transient A short (momentary) impulse.

Transition Band The transition band begins at the nominal cutoff frequency, until the stop band. See Chapter 23.

Transparent A device which is transparent sounds as clean on its output as the source.

True Peak Level An interpolated measurement of peaks that are occurring in converters or filters, defined by ITU BS.1770. Embrace the true peak level! It will keep you out of trouble. See Chapter 16.

Truncation Reduction of wordlength by cutting off the lower bits. If dithering was not performed first, then simple wordlength truncation causes distortion.

U

Uncompressor Upward expander. It is not true that we can remove the distortion or artifacts caused by excessive compression with an upward expander, but we can mitigate some of the damage with judicious and careful thresholds, attack and release time constants. See Chapter 7.

Unity Gain The ratio of output to input level is 1, or 0 dB.

Upsampling Raising the sample rate of material so as to reduce distortion by nonlinear digital processes such as compression. See Chapter 9.

Upward Compression Raising the level of low passages. See Chapter 7.

Upward Expansion takes high-level passages and brings them up even further. See Chapter 7.

V

Variable Mu (Vari-Mu) A type of audio tube whose gain is adjustable by a DC control voltage. Manley has trademarked a variation on the name: "Vari-Mu".

VCA (Voltage Controlled Amplifier), is an audio amplifier whose gain is controlled by a control voltage.

Volume Formally: A measure of the capacity of a container, e.g. quarts, liters, cubic meters. This is not a formal acoustical term. When used informally by audio people, it has three possible meanings: gain, level or loudness, so its use is ambiguous. In formal discussion, it is suggested to use gain, level or loudness instead. See Chapter 16.

W

Weighting Altering the frequency response so as to de-emphasize ranges where the human ear is less sensitive and emphasize ranges where the ear is more sensitive. Used when measuring signal to noise ratio in a psychoacoustic manner.

Wordlength Also known as **Bit Depth**. The number of discrete data bits employed to transfer a digital value from the source to a destination.

Wow and Flutter Speed variations in recordings, commonly caused by imperfections in analog tape recorders.

Z

Zip A file format that is popular for its data compression and archival purposes, to enable reliable and faster up and downloads. ZIP can store one or more files in a single archive.

Index

T

U