INTRODUCTION TO DIGITAL AUDIO CODING AND STANDARDS

THE KLUWER INTERNATIONAL SERIES
IN ENGINEERING AND COMPUTER SCIENCE

INTRODUCTION TO DIGITAL AUDIO CODING AND STANDARDS

Marina Bosi
Stanford University

Richard E. Goldberg
The Brattle Group

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

Bosi, Marina
Introduction to Digital Audio Coding and Standards / by Marina Bosi, Richard E. Goldberg.
p. cm. –(The Kluwer international series in engineering and computer science; SECS 721
Includes bibliographical references and index.
ISBN 978-1-4613-5022-4 ISBN 978-1-4615-0327-9 (eBook)
DOI 10.1007/978-1-4615-0327-9
1. Sound—Recording and reproducing—Digital techniques—Standards. 2. Data compression (Computer science) I. Goldberg, Richard E. II. Title. III. Series.

TK7881.4 .B59 2002
621.389'32—dc21

2002040686

Originally published by Kluwer Academic Publishers in 2003
Softcover reprint of the hardcover 1st edition 2003

Printed on acid-free paper.

Contents

About the Authors

Marina Bosi is a Consulting Professor at Stanford University's Computer Center for Research in Music and Acoustics (CCRMA) and Chief Technology Officer of MPEG LA®, a firm specializing in the licensing of multimedia technology. Past president of the Audio Engineering Society, Dr. Bosi is the author of numerous articles and the holder of several patents in audio technology. Dr. Bosi has been involved in the development of MPEG, Dolby, and DTS audio coders.

Richard E. Goldberg is a Partner at The Brattle Group, a management consulting firm specializing in economics and finance issues. Dr. Goldberg's practice focuses on business valuation and risk management. Dr. Goldberg has a Ph.D. in Physics from Stanford University and an A.B. in Astrophysics from Princeton University. Audio coding technology and related business applications have long been areas of interest for him.

Foreword

THE RISE OF DIGITAL AUDIO

Leonardo Chiariglione – Telecom Italia Lab, Italy

Analogue speech in electrical form has a history going back more than a century and a quarter to the early days of the telephone. However, interest in digital speech only gathered momentum some 40 years ago when the telecommunications industry started a global project to digitize the telephone network. The technology trade-off of the time in this infrastructure-driven project led to a preference for adding transmission capacity over finding methods to reduce the bitrate of the speech signal so the use of compression technology for speech remained largely dormant. When in the late 1980s the ITU-T standard for visual telephony became available enabling compression of video by a factor of 3,000, the only audio format in use to accompany this highly compressed video was standard telephone quality 64 kb/s PCM. It was only where transmission capacity was a scarce asset, like in the access portion of radiotelephony, that speech compression became a useful tool.

Analogue sound in electrical form has a history going back only slightly more than a century ago when a recording industry began to spring up around the gramophone and other early phonographs. The older among us fondly remember collections of long playing records (LPs) which later gave way to cassette tapes as the primary media for analogue consumer audio. Interest in digital audio received a boost some 20 years ago when the

Consumer Electronics (CE) industry developed a new digital audio recording medium: a 12 cm platter – the compact disc (CD) – carrying the equivalent of 70 minutes of uncompressed stereo digital audio. This equivalent of one long playing (LP) record was all that the CE industry needed at the time and compression was disregarded as the audio industry digitized.

Setting aside some company and consortium initiatives, it was only with the MPEG-1 project in the late 1980s that compressed digital audio came to the stage. MPEG-1 had the ambitious target of developing a single standard addressing multiple application domains: the digital version of the old compact cassette, digital audio broadcasting, audio accompanying digital video in interactive applications, the audio component of digital television and professional applications were listed as the most important.

The complexity of the task was augmented by the fact that each of these applications was targeted to specific industries and sectors of those industries, each with their own concerns when it comes to converting a technology into a product. The digital version of the old compact cassette was the most demanding: quality of compressed audio had to be good, but the device had to be cheap; in digital audio broadcasting quality was at premium, but the device had to have an affordable price; audio in interactive audio-visual applications could rely on an anticipated mass market where a high level of silicon integration of all decompression functionalities could be achieved; a similar target existed for audio in digital television; lastly, many professional applications required the best quality possible at the lowest possible bitrates.

It could be anticipated that these conflicting requirements would make the task arduous, and indeed the task turned out to be so. But the Audio group of MPEG, in addition to being highly competitive, was also inventive. Without calling them so, the Audio group was the first to define what are now known as "profiles" under the name of "layers". And quite good profiles they turned out to be because a Layer I bitstream could be decoded by a Layer II and a Layer III decoder in addition to its own decoder, and a Layer II bitstream could be decoded by a Layer III decoder in addition to its own decoder.

The MPEG-2 Audio project later targeted multichannel audio, but the story was a complicated one. With MPEG-1 Audio providing transparent quality at 256 kb/s for a stereo signal with Layer II coding and the same quality at 192 kb/s with Layer III coding, it looked like a natural choice that MPEG-2 Audio should be backwards compatible, in the sense that an MPEG-1 Audio decoder of a given layer should be able to decode the stereo component of an MPEG-2 Audio bitstream. But it is a well-known fact that backwards compatible coding provides substantially lower quality compared to unconstrained coding. This was the origin of the bifurcation of the multichannel audio coding work: Part 3 of MPEG-2 specifies a backward compatible multichannel audio coding and Part 7 of MPEG-2 (called Advanced Audio Coding – AAC) a non backward compatible or unconstrained multichannel audio coding standard.

AAC has been a major achievement. In less than 5 years after approving MPEG-1 Audio layer III, the MPEG Audio group produced an audio compression standard that offered transparency of stereo audio down to 128 kb/s.

This book has been written by the very person who led the MPEG-2 AAC development. It covers a gap that existed so far by offering both precious information on digital audio in general and in-depth information on the principles and practice of the 3 audio coding standards MPEG-1, MPEG-2 and MPEG-4. Its reading is a must for all those who want to know more, for curiosity or professional needs, about audio compression, a technology that has led mankind to a new relationship with the media.

Preface

The idea of this book came from creating and teaching a class for graduate students on Audio Coding at Stanford University's Computer Center for Research in Music and Acoustics (CCRMA). The subject of audio coding is a "hot topic" with students wanting to better understand the technology behind the MP3 files they are downloading over the internet, their audio choices on their DVDs, the digital radio proposals in the news, and the digital television offered by cable and satellite providers. Now in its sixth year, the class attracts a wide range of participants including music students, engineering students, and industrial professionals working in telecommunications, hardware design, and software product development.

In designing a course for such a diverse group, it is important to develop a shared vocabulary and understanding of the basic building blocks of a digital audio coder so that the choices made in any particular coder can be discussed using a commonly understood language. In the course, we first address the theory and implementation of each of the basic coder building blocks. We then show how the building blocks fit together into a full coder and how to judge the performance of such a coder. Finally, we discuss the features, choices, and performance of the main state-of-the-art coders in commercial use today.

The ultimate goal of the class, and now of this book, is to present the student and the reader with a solid enough understanding of the major issues in the theory and implementation of perceptual audio coders that they are able to build their own simple audio codec. MB is always very pleasantly surprised to hear the results of her student's work. As a final project for the class, they are able to design and implement perceptual audio coding schemes equivalent to audio coding schemes that were state-of-the-art only a few years ago. It is our hope that this book will allow advanced readers to achieve similar goals.

The book is organized in two parts: The first part consists of Chapters 1 through 10 which present the student with the theory of the major building blocks needed to understand the workings of a perceptual audio coder. The second part consists of Chapters 11 through 15 in which the most widely used perceptual audio coders are presented and their major features discussed. Typically, the students start their final project (building their own perceptual audio coder) at the transition from the first part to the second. In this manner, they are confronting their own trade-offs in coder design while hearing how these very same trade-offs are handled in state-of-the-art commercial coders. The particular chapter contents are as follows:

Chapter 1 serves as an introductory chapter in which the goals and high-level structure of audio coders are discussed.

Chapter 2 discusses how to quantize sampled data so that it can be represented with a finite number of bits for storage or transmission. Errors introduced in the quantization process are discussed and compared for uniform and floating point quantization schemes. The ideas of noiseless (entropy) coding and Huffman coding are introduced as means for further reducing the bit requirement for quantized data.

Chapter 3 addresses sampling in the time domain and how to later recover the original continuous time input signal. The basics of representing audio signals in the frequency domain via Fourier Transforms are also introduced.

Chapters 4 and 5 present the main filter banks used for implementing the time to frequency mapping of audio signals. Quadrature Mirror filters and their generalizations, Discrete Fourier Transforms, and transforms based on Time Domain Aliasing Cancellation are all analyzed. In addition, methods for designing time variant filter banks are illustrated.

Chapters 6 addresses the fundamentals of psychoacoustics and human hearing. Chapter 7 then discusses applications of frequency and temporal masking effects to develop masking curves for use in audio coding.

Chapter 8 presents methods for allocating bits to differing frequency components so as to maximize audio quality at a given bitrate. This chapter shows how the masking curves discussed in the previous chapter can be exploited to reduce audio coding bitrate.

Chapter 9 discusses how the pieces described in the previous chapters fit together to create a perceptual audio coding system. The standardization process for audio coders is also discussed.

Chapter 10 is devoted to the understanding of methods for evaluating the quality of audio coders.

Chapter 11 gives an overview MPEG-1 Audio. The different audio layers are discussed as well implementation and performance issues. MPEG Layer III is the coding scheme used to create the well-known MP3 files.

Chapters 12 and 13 present the second phase of MPEG Audio, MPEG-2, extending the MPEG-1 functionality to multichannel coding, to lower sampling frequencies, and to higher quality audio. MPEG-2 LSF, MPEG-2 BC, and MPEG-2 AAC are described. The basics of multichannel and binaural coding are also introduced in these chapters.

Chapter 14 is devoted to Dolby AC-3, the audio coder used in digital television standards and in DVDs.

Chapter 15 introduces the latest MPEG family of audio coding standards, MPEG-4, which allows for audio coding at very low bit rates and other advanced functionalities. MPEG-4 looks to be the coding candidate of choice for deployment in emerging wireless and wired network applications.

Acknowledgements

Audio coding is an area full of lore where you mostly learn via shared exploration with colleagues and the generous sharing of experience by previous explorers. This book is our attempt to pass on what we've learned to future trekkers. Some of the individuals we have been lucky enough to learn from and with during our personal explorations include: Louis Fielder and Grant Davidson from Dolby Laboratories; Karlheinz Brandenburg, Martin Dietz, and Jürgen Herre from the Fraunhofer Gesellschaft; Jim Johnston and Schulyer Quackenbush from AT&T; Leonardo Chariglione the esteemed MPEG Convener; Gerhard Stoll from IRT; and David Mears from the BBC. To all of the above (and the many others we've had the privilege to work with), we offer heartfelt thanks for their generosity of spirit and shared good times.

The course this book is based upon came into being due to the encouragement of John Chowning, Max Mathews, Chris Chafe, and Julius Smith at Stanford University. It was greatly improved by the nurturing efforts of its über-TA Craig Sapp (whose contributions permeate the course, especially the problem sets) and the feedback and good sportsmanship of its many students over the last 6 years. Thanks to Louis Fielder, Dan Slusser of DTS, and Baryn Futa of MPEG LA® for allowing MB to fit teaching into a full-time work schedule. Thanks also to Karlheinz Brandenburg and Louis Fielder for their guest lectures on MP3 and the Dolby coders, respectively,

and to Louis for hosting the class at the Dolby facilities to carry out listening tests.

Not being able to find an appropriate textbook, the course made due for several years with extensive lecture notes. That would probably still be the case were it not for the intervention and encouragement of Joan L. Mitchell, IBM Fellow. Joan made the writing of a book seem possible and shared her hard-won insight into the process. You would not be holding this book in your hands were it not for Joan's kind but forceful encouragement.

Thanks to Joan Mitchell, Bernd Edler from Universität Hannover, Leonardo Chariglione, Louis Fielder, and Karlheinz Brandenburg for their careful review of early drafts of this book – their comments and feedback helped the clarity of presentation immensely. Thanks to Sarah Kane of the Brattle Group for her tireless yet cheerful administrative support during the writing process. Thanks also to Baryn Futa and Jamie Read from the Brattle Group for their support in ensuring that work demands didn't prevent finding the time for writing.

In spite of the generous help and support of many individuals, there are surely still some murky passages and possibly errors in the text. For any such blemishes, the authors accept full responsibility. We do sincerely hope, however, that you find enough things of novelty and beauty in the text that any such findings seem minor in comparison.

To Alex