McGRAW-HILL VIDEO/AUDIO PROFESSIONAL

Principles of Digital Audio

FOURTH EDITION

KEN C. POHLMANN

Principles of Digital Audio

Ken C. Pohlmann

Fourth Edition

McGraw-Hill

New York San Francisco Washington, D.C. Auckland Bogotá
Caracas Lisbon London Madrid Mexico City Milan
Montreal New Delhi San Juan Singapore
Sydney Tokyo Toronto

Library of Congress Cataloging-in-Publication Data

Pohlmann, Ken C.
Principles of digital audio / Ken C. Pohlmann.—4th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-07-134819-0
1. Sound—Recording and reproducing—Digital techniques. I. Title.
TK7881.4 P63 2000
621.389'3—dc21
99-054165

McGraw-Hill

A Division of The **McGraw-Hill** *Companies*

1 2 3 4 5 6 7 8 9 0 AGM/AGM 9 0 5 4 3 2 1 0

ISBN 0-07-134819-0

The sponsoring editor for this book was Stephen S. Chapman and the production supervisor was Maureen Harper. It was set in Century Schoolbook by Pro-Image Corporation.

Printed and bound by Quebecor/Martinsburg.

This book was printed on recycled, acid-free paper containing a minimum of 50% recycled, de-inked fiber.

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please write to the Director of Special Sales, Professional Publishing, McGraw-Hill, Two Penn Plaza, New York, NY 10121-2298. Or contact your local bookstore.

the modulated lapped transform (MLT). In the MDCT, the length of the overlapping windows is twice that of the block time (shift length of the transform). Frequency-domain subsampling is performed; the number of time and frequency components equals the shift length of the input time-domain sampled signal. MDCT also lends itself to adaptive window switching approaches with different window functions for the first and second half of the window; the time domain aliasing property must be independently valid for each window half. Many bands are possible with the MDCT with good efficiency, on the order of an FFT computation. Many codecs apply a window function to blocks prior to transformation; this helps minimize spectral leakage of spectral coefficients. A window is a time function that is multiplied by an audio block to provide a windowed audio block; the window shape governs the frequency selectivity of the filter bank. The overlap/add characteristic minimizes blocking artifacts and leakage of spectral coefficients. Digital filters and windows are discussed in chapter 17.

Hybrid filter banks use a cascade of different filter types (such as polyphase and MDCT) to provide different frequency resolutions at different frequencies with moderate complexity; for example, MPEG-1 Layer III encoders use a hybrid filter with a polyphase filter bank and MDCT. The ATRAC algorithm used in the MiniDisc, examined in chapter 12, is a hybrid coder that uses QMF to divide the signal into three subbands, then each subband is transformed into the frequency domain using the MDCT. Table 10.3 compares the properties of filter banks used in several low-bit rate coders.

MPEG-1 Audio Standard

The International Standards Organization and the International Electrotechnical Commission formed the Moving Pictures Expert Group (MPEG) in 1988 to devise compression techniques for audio and video. This group has

TABLE 10.3 Comparison of filter-bank properties.

Feature	Layer 1	Layer 2	Layer 3	AC-2	AC-3	ATRAC*	PAC/MPAC
Filterbank type	PQMF	PQMF	Hybrid PQMF/ MDCT	MDCT/MDST	MDCT	Hybrid QMF/ MDCT	MDCT
Frequency resolution at 48 kHz	750 Hz	750 Hz	41.66 Hz	93.75 Hz	93.75 Hz	46.87 Hz	23.44 Hz
Time resolution at 48 kHz	0.66 ms	0.66 ms	4 ms	1.3 ms	2.66 ms	1.3 ms	2.66 ms
Impulse response (LW)	512	512	1664	512	512	1024	2048
Impulse response (SW)	—	—	896	128	256	128	256
Frame length at 48 kHz	8 ms	24 ms	24 ms	32 ms	32 ms	10.66 ms	23 ms

*ATRAC is operating at a sampling frequency at 44.1 kHz. For comparison, the frame length and impulse response figures are given for an ATRAC system working at 48 kHz.
(*Brandenburg and Bosi*)

developed several highly successful standards. It first devised the ISO/IEC International Standard 11172 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s" for reduced data rate coding of digital video and audio signals; the standard was finalized in November, 1992. It is commonly known as MPEG-1. (The acronym is pronounced "m-peg.") The standard has three major parts: system (multiplexed video and audio), video, and audio; a fourth part defines conformance testing. The maximum audio bit rate is set at 1.856 Mbps. The audio portion of the standard (11172-3) has found many applications such as Video CD, CD-ROM, ISDN, video games, and digital audio broadcasting. It supports coding of 32, 44.1, and 48 kHz PCM data at bit rates of approximately 32 to 224 kbps/channel (64 to 448 kbps for stereo). (Because data networks use data rates of 64 kbps (8 bits sampled at 8 kHz), most coders output a data channel rate that is a multiple of 64.)

The ISO/MPEG-1 standard was specifically developed to support audio and video coding for CD playback within the CD's bandwidth of 1.41 Mbps. However, the standard supports stereo bit rates ranging from 64 kbps to 448 kbps, as well as mono audio coding of 32 kbps. In addition, in the stereo modes, stereophonic irrelevance and redundancy can be optionally exploited to reduce the bit rate. Stereo audio bit rates below 256 kbps are useful for applications requiring more than two audio channels while maintaining full screen motion video. Rates above 256 kbps are useful for applications requiring higher audio quality, and partial screen video images. In either case, the bit allocation is dynamically adaptable according to need. The MPEG-1 standard is based on a history of research and development of data reduction algorithms.

MUSICAM (Masking-pattern Universal Subband Integrated Coding And Multiplexing) was an early and successful perceptual coding algorithm. Derived from MASCAM (Masking-pattern Adapted Subband Coding And Multiplexing), MUSICAM divides the input audio signal into 32 subbands and uses perceptual coding models of minimum hearing threshold and masking to achieve data reduction. With a sampling frequency of 48 kHz, the subbands are each 750 Hz wide. Each subband is given a 6-bit scale factor according to the peak value in the subband's 12 samples and quantized with a variable word ranging from 0 to 15 bits. Scale factors are calculated over a 24-ms interval, corresponding to 36 samples. A subband is quantized only if it contains audible signals above the masking threshold. Subbands with signals well above the threshold are coded with more bits, yielding a higher S/N ratio. In other words, within a given bit rate, bits are assigned where they are most needed. In addition, a side-chain Fourier spectral analysis is performed on the input signal to assist in the masking threshold calculations. In this way, the data rate is reduced, to perhaps 128 kbps per mono channel (256 kbps for stereo). Extensive tests of 128 kbps MUSICAM showed that the coder achieves fidelity that is indistinguishable from a CD source, that it is monophonically compatible, that at least two cascaded codec stages produce no audible degradation, and that it is preferred to very high quality FM signals.

The audio portion of the ISO/MPEG-1 standard can trace its origins to tests conducted by Swedish Radio in July 1990. MUSICAM coding was judged su-

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.