Preface

Since branching out of speech coding in the early 1970s, audio coding has now slipped into our daily lives in a variety of applications, such as mobile music/video players, digital television/audio broadcasting, optical discs, online media streaming, and electronic games. It has become one of the essential technologies in today’s consumer electronics and broadcasting equipments.

In its more than 30 years of evolution, many audio coding technologies had come into the spotlight and then became obsolete, but only a minority have survived and are deployed in major modern audio coding algorithms. While covering all the major turns and branches of this evolution is valuable for technology historians or for people with intense interests, it is distracting and even inundating for most readers. Therefore, those historic events will be omitted and this book will, instead, focus on the current state of this evolution. Such a focus also helps to provide full coverage to selected topics in this book.

This state of the art is presented from the perspective of a practicing engineer and adjunct associate professor, who single-handedly developed the whole DRA audio coding standard, from algorithm architecture to assembly-code implementation and to subjective listening tests. This perspective has a clear focus on “why” and “how to.” In particular, many purely theoretical details such as proof of perfect reconstruction property of various filter banks are omitted. Instead, the emphasis is on the motivation for a particular technology, why it is useful, what it is, and how it is integrated into a complete algorithm and implemented in practical products. Consequently, many practical aspects of audio coding technologies normally excluded in audio coding books, such as transient detection and implementation of decoders on low-cost microprocessors, are covered in this book.

This book should help readers to grasp the state-of-the-art audio coding technologies and build a solid foundation for them to either understand and implement various audio coding standards or develop their own should the need arise. It is, therefore, a valuable reference for engineers in the consumer electronics and broadcasting industry and for graduate students of electrical engineering.

Audio coding seeks to achieve data compression by removing perceptual irrelevance and statistical redundancy from a source audio signal and the removal efficiency is powerfully augmented by data modeling which compacts and/or decorrelates the
source signal. Therefore, the presentation of this book is centered around these three basic elements and organized into the following five parts.

Part I gives an overview of audio coding, describing the basic ideas, the key challenges, important issues, fundamental approaches, and the basic codec architecture.

Part II is devoted to quantization, the tool for removing perceptual irrelevancy. Chapter 2 delineates scalar quantization which quantizes a source signal one sample at a time. Both uniform and nonuniform quantization, including the Lloyd–Max algorithm, are discussed. Companding is posed as a structured and simple method to implement nonuniform quantization.

Chapter 3 describes vector quantization which quantizes two or more samples of a source signal as one block each time. Also included is the Linde–Buzo–Gray (LBG) or k-means algorithm which builds an optimal VQ codebook from a set of training data.

Part III is devoted to data modeling which transforms a source signal into a representation that is energy-compact and/or decorrelated. Chapter 4 describes linear prediction which uses a linear combination of the historic samples of the source signal as a prediction for the current sample so as to arrive at a prediction error signal that has lower energy and is decorrelated. It first explains why quantizing the prediction error signal, instead of the source signal, can dramatically improve coding efficiency. It then presents open-loop DPCM and DPCM, the two most common forms of linear prediction, derives the normal equation for optimal prediction, presents Levinson–Durbin algorithm that iteratively solves the normal equation, shows that the prediction error signal has a white spectrum and is thus decorrelated, and illustrates that the prediction decoder filter provides an estimate of the spectrum of the source signal. Finally, a general framework for linear prediction that can shape the spectrum of quantization noise to desirable shapes, such as that of the absolute threshold of hearing, is presented.

Chapter 5 deals with transforms which linearly transform a block of source signal samples into another block of coefficients whose energy is compacted to a minority. It first explains why this compaction of energy leads to dramatically improved coding efficiency through the AM–GM inequality and the associated optimal bit allocation strategy. It then derives the Karhunen–Loeve transform from the search for the optimal transform. Finally, it presents suboptimal and practical transforms, such as discrete Fourier transform (DFT) and discrete cosine transform (DCT).

Chapter 6 presents subband filter banks as extended transforms in which historic blocks of source samples overlap with the current block. It describes various aspects of subband coding, including reconstruction error and polyphase representation and illustrates that the dramatically improved coding efficiency is also achieved through energy compaction.

Chapter 7 is devoted to cosine modulated filter banks (CMFB), whose structure is amenable for fast implementation. It first builds this filter bank from DFT and explains that it has a structure of a prototype filter plus cosine modulation. It then presents nonperfect reconstruction and perfect reconstruction CMFB and their efficient implementation structures. Finally, it illustrates that modified discrete cosine transform (MDCT), the most widely used filter bank in audio coding, is a special and simple case of CMFB.
Part IV is devoted to entropy coding, the tool for removing statistical redundancy. Chapter 8 establishes that entropy is determined by the probability distribution of the source signal and is the fundamental lower limit of bit rate reduction. It then shows that any meaningful entropy codes have to be uniquely decodable and, to be practically implementable, should be instantaneously decodable. Finally, it illustrates that prefix-free codes are just such codes and further proves Shannon’s noiseless coding theorem, which essentially states that the entropy can be asymptotically approached by a prefix-free code if source symbols are coded as blocks and the block size goes to infinity.

Chapter 9 presents Huffman code, an optimal prefix-free code widely used in audio coding. It first presents Huffman’s algorithm, which is an iterative procedure to build a prefix-free code from the probability distribution of the source signal, and then proves its optimality. It also addresses some practical issues related to the application of Huffman coding, emphasizing the importance of coding source symbols as longer blocks.

While the previous parts can be applied to signal coding in general, Part V is devoted to audio. Chapter 10 covers perceptual models which determine which part of the source signal is inaudible (perceptually irrelevant) and thus can be removed. It starts with the absolute threshold of hearing, which is the absolute sensitivity level of the human ear. It then illustrates that the human ear processes audio signals in the frequency domain using nonlinear and analog subband filters and presents Bark scale and critical bands as tools to describe the nonuniform bandwidth of these subband filters. Next, it covers masking effects which describe the phenomenon that a weak sound becomes less audible due to the presence of a strong sound nearby. Both simultaneous and temporal masking are covered, but emphasis is given to the former because it is more thoroughly studied and extensively used in audio coding. The rest of the chapter addresses a few practical issues, such as perceptual bit allocation, converting masked threshold to the subband domain, perceptual entropy, and an example perceptual model.

Chapter 11 addresses the resolution challenge posed by transients. It first illustrates that audio signals are mostly quasistationary, hence need fine frequency resolution to maximize energy compaction but are frequently interrupted by transients, which requires fine time resolution to avoid “pre-echo” artifacts. The challenge, therefore, arises: a filter bank cannot have fine frequency and time resolution simultaneously according to the Fourier uncertainty principle. It then states that one approach to address this challenge is to adapt frequency resolution in time to the presence and absence of transients and further presents switched-window MDCT as an embodiment: switching the window length of MDCT in such a way that short windows are applied to transients and long ones to quasistationary episodes. Two such examples are given, which can switch between two and three window lengths, respectively. For the double window length example, two more techniques, temporal noise shaping and transient localization are given, which can further improve the temporal resolution of the short windows. Practical methods for transient detection are finally presented.
Chapter 12 deals with joint channel coding. Only two widely used methods are covered, they are joint intensity coding and sum/difference (M/S stereo) coding. Methods to deal with low-frequency effect (LFE) channels are also included.

Chapter 13 covers a few practical issues frequently encountered in the development of audio coding algorithms, such as how to organize various data, how to assign entropy codebooks, how to optimally allocate bit resources, how to organize bits representing various compressed data and control commands into a bit stream suitable for transmission over various channels, and how to make the algorithm amenable for implementation on low-cost microprocessors.

Chapter 14 is devoted to performance assessment, which, for a given bit rate, becomes an issue of how to evaluate coding impairments. It first points out that objective methods are highly desired, but are generally inadequate, so subjective listening tests are necessary. The double-blind principle of subjective listening test is then presented, along with the two methods, namely the ABX test and ITU-R BS.1116, that implement it.

Finally, Chap. 15 presents Dynamic Resolution Adaptation (DRA) audio coding standard as an example to illustrate how integrate the technologies described in this book to create a practical audio coding algorithm. DRA algorithm has been approved by the Blu-ray Disc Association as part of its BD-ROM 2.3 specification and by Chinese government as its national standard.

Yuli You
Adjunct Associate Professor
Department of Electrical and Computer Engineering
yuliyou@hotmail.com
Audio Coding
Theory and Applications
You, Y.
2010, XVI, 344 p., Hardcover