

Biorthogonality in Lapped Transforms: A Study in High-Quality Audio Compression

by

Shiufun Cheung

E.E., Massachusetts Institute of Technology (1993)

S.M., Massachusetts Institute of Technology (1991)

S.B., Massachusetts Institute of Technology (1989)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

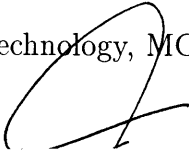
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1996

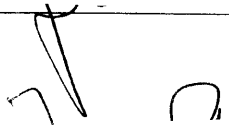
© Massachusetts Institute of Technology, MCMXCVI. All rights reserved.

Author


Department of Electrical Engineering and Computer Science

January 31, 1996

Certified by


Jae S. Lim
Professor of Electrical Engineering
Thesis Supervisor

Accepted by


Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

APR 11 1996

ENG.

LIBRARIES

**Biorthogonality in Lapped Transforms:
A Study in High-Quality Audio Compression**

by
Shiufun Cheung

Submitted to the Department of Electrical Engineering and Computer Science
on January 31, 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

The demand for high-quality audio in transmission systems such as Digital Audio Broadcast (DAB) and High-Definition TeleVision (HDTV), as well as commercial products such as the MiniDisc (MD) and the Digital Compact Cassette (DCC), has generated considerable interest in audio compression schemes. The common objective is to achieve high quality at a rate significantly smaller than the 16 bits/sample used in current Compact Disc (CD) and Digital Audio Tape (DAT) systems. This thesis explores the current state of audio compression research, placing special emphasis on one important aspect: the short-time spectral decomposition.

In conventional audio coders, the short-time spectral decomposition serves to recast the audio signal in a representation that is not only amenable to perceptual modeling but also conducive to deriving transform coding gain. This decomposition is commonly achieved by a multirate filter bank, or equivalently, a lapped transform.

Towards the goal of improving the performance of audio compression schemes, this thesis contains the formulation of a biorthogonal cosine-modulated filter bank which is a generalization of Malvar's extended lapped transform (ELT). The ELT, a popular implementation of cosine-modulated filter banks, is of particular interest because it forms the major building block of signal decomposition schemes in many audio coders.

Conventional lapped transforms are designed to be orthogonal filter banks in which the analysis and synthesis filters are identical. Allowing the analysis and synthesis filters to differ leads to a biorthogonal transform which has more degrees of design freedom than its orthogonal counterpart. The thesis contains proofs for three special cases and for the general case that the incorporation of biorthogonality into an M -channel ELT yields an increase of $M/2$ degrees of freedom. This additional flexibility allows the design of synthesis filter banks with improved sidelobe behavior which should be beneficial to audio coder performance.

Thesis Supervisor: Jae S. Lim
Title: Professor of Electrical Engineering

Dedication

to

Josephine

*My Beloved,
Who Never Doubted.*

Acknowledgements

In my quixotic pursuit of a doctorate degree, I found that acquisition of knowledge only heightens my awareness of my own ignorance and inadequacy. The completion of this dissertation would not have been possible had it not been for the assistance and support, technical or otherwise, from many individuals.

Professor Jae S. Lim supervised my academic research for the last seven years. I will never cease to be amazed by the limitless energy and dazzling efficiency he exhibits in his numerous endeavors. For the opportunities he has provided me throughout my doctoral program, I am eternally in his debt.

Other members of the thesis committee have also been extremely supportive. Professor Gilbert Strang has the uncanny ability of seeing into the heart of every problem I encountered in my formulations. His mathematical insight directed me to the path that eventually led to the central theorem in the dissertation. Professor David Staelin has been a constant presence in my career at MIT. He supervised my bachelor thesis, sat on my area examination committee and graciously consented to be a reader on short notice. The thesis committee have been especially accommodating in the scheduling of the defense during their sabbatical periods. For this and many other instances of kindness, I am thankful.

I am also fortunate enough to have conducted brief but enlightening conversations with some distinguished researchers in the fields of digital signal processing and digital audio compression. Among them are Dr. Karlheinz Brandenburg, Dr. Henrique Malvar and Professor Hamid Nawab. Discussions with them substantially enhanced the quality of this dissertation.

Of the members of the Advanced Television and Signal Processing (ATSP) group, past and present, a few are extremely brilliant, some offer invaluable advice, many are perceptive and insightful teachers, all I am honored to count as my friends. Within this diverse group, the following individuals deserve special recognition. John Apostolopoulos, a model of sincere scholarship and disarming courtesy, has always kept me abreast of developments in related research efforts. His assistance has indirectly contributed a lot to my studies. Babak Ayazifar, from whom I learned the art of sparring wit and sarcastic humor, tried without success to provide me with research and business opportunities. His efforts have nevertheless been much appreciated. Peter Monta, for whom the label “genius” is not an exaggeration, taught me the practical aspects of electronics and computer programming. We worked together on the MIT Audio Coder project and I looked upon him as a mentor. Lon Sunshine has been my companion in my transitory bachelorhood. We shared not only an office but also common hobbies and interests. This list would not be complete without mentioning the devoted secretaries of the research group, which include Debra Martin and Cindy LeBlanc. Cindy in particular deserves special mention. Generous of spirit and incredibly

Acknowledgements

competent, she has been a constant source of encouragement and support.

Many others, from MIT and beyond, have helped to make my student life a memorable and pleasant one. Among them, I would like first to single out Ms. Lillian Whelpley, my international host. She is my parent away from home and I look forward to many more of her mouth-watering feasts. Mattan Kamon, my apartment mate for the last few years, is an immensely popular man who has enlarged my social circle considerably. Many of my now-very-close friends, such as Chavela Carr and Joel Phillips, I know through him. Chi-yuan Hsu, my best friend in high school, miraculously ended up practicing medicine in Boston. I am definitely one of those who benefitted from his medical expertise and his wisdom. Helen Meng and Warren Lam complete the circle of my friends from Hong Kong. It is always comforting to be able to converse in one's mother tongue. Last but not least, I owe a lot of my social development to one whom I hold dear, Bradley Gwynn. He opened my eyes to an alternative lifestyle and for this I can never repay him.

My parents, who never understood much of this, have waited long for me to finish. I thank them for their patience. My brother, Shiu Kei, has been my advisor in family matters and in career decisions. While I wish him every success in his new job in Hong Kong, I will sorely miss his counsel.

To Josephine, trusted confidant, loyal comrade and my chosen mate, I dedicate this work. Her presence alone is a nepenthe in trying times. I simply cannot find the words to express adequately my love and gratitude.

The road to the doctorate degree is a long and arduous one. Many friends and colleagues offered moral support along the way. To those I have inadvertently neglected to mention, I apologize and extend to you my heartfelt thanks.

Contents

1	Introduction	12
1.1	Context & Motivation	12
1.1.1	Rise of Digital Audio	12
1.1.2	Need for Audio Compression	13
1.1.3	Audio Compression & High-Definition Television	13
1.1.4	Multidimensional & Multichannel Audio	13
1.2	Dissertation	14
1.2.1	Scope of Thesis	14
1.2.2	Organization of Thesis	15
2	Digital Audio Compression	16
2.1	Psychoacoustics	16
2.1.1	Critical Band Analysis	17
2.1.2	Threshold in Quiet	17
2.1.3	Masking Effects	19
2.1.3.1	Simultaneous (Spectral) Masking	20
2.1.3.2	Non-Simultaneous (Temporal) Masking	20
2.2	Frequency-Domain Coding of Audio	20
2.2.1	Short-Time Spectral Decomposition	21
2.2.2	Perceptual Modeling	22
2.2.3	Quantization and Codeword Assignment	23
2.3	Other Audio Coding Schemes	24

2.3.1	Linear Predictive Coders	24
2.3.2	Subband ADPCM Coders	24
2.3.3	Dynamic Dictionary	25
2.4	Multichannel Audio Compression	25
2.4.1	Sum-Difference Stereo & Matrix Surround	25
2.4.2	Intensity Stereo & Intensity Surround	25
2.5	Major Research & Development Efforts in Audio Compression	26
2.5.1	Erlangen, Germany	26
2.5.1.1	Low-Complexity Adaptive Transform Coding (LC-ATC)	26
2.5.1.2	Optimum Coding in the Frequency Domain (OCF)	26
2.5.2	AT&T Bell Laboratories	27
2.5.3	Centre National d'Etudes des Télécommunications (CNET)	27
2.5.4	Adaptive Spectral Entropy Coding (ASPEC)	28
2.5.5	Masking Pattern Adapted Universal Subband Integrated Coding and Multi- plexing (MUSICAM)	28
2.5.6	Moving Pictures Expert Group (MPEG)	28
2.5.7	Digital Compact Cassette (DCC) & MiniDisc (MD)	29
2.5.8	Dolby Laboratories	29
2.5.9	MIT-Advanced Television and Signal Processing Group (ATSP)	30
2.6	Parameters for Assessing Audio Coders	30
2.6.1	Subjective Quality	30
2.6.2	Objective Measurements	31
2.6.3	Bit Rate, System Delay & Complexity	31
3	Multirate Filter Banks & Audio Signal Representations	32
3.1	Theory of Multirate Filter Banks	32
3.1.1	Basics	32

3.1.2	Polyphase Representation	33
3.1.3	Block Transform	34
3.1.4	Time-Frequency Tiling & Complete Representations	35
3.2	Multirate Filter Banks in Audio Compression	36
3.2.1	From Disjoint to Overlapping Temporal Frames	36
3.2.2	From Uniform to Nonuniform Filter Banks	37
3.2.3	The Emergence of Time-Varying Filter Banks	40
4	Biorthogonality in Lapped Transforms	42
4.1	Lapped Transforms	42
4.1.1	Extended Lapped Transforms & Cosine-Modulated Filter Banks	44
4.1.2	Perfect Reconstruction in the Extended Lapped Transform	45
4.2	Generalization of the Extended Lapped Transform	45
4.2.1	Time Domain Analysis	45
4.2.2	Simplification of the Constraints	48
5	Advantages from Incorporation of Biorthogonality	51
5.1	Increase in Degrees of Freedom: Special Cases	51
5.1.1	Special Case I: $K = 1$	51
5.1.2	Special Case II: $K = 2$	54
5.1.3	Special Case III: $K = 3$	57
5.2	Increase in Degrees of Freedom: General Case	60
5.2.1	Emerging Pattern from Special Cases	60
5.2.2	Theorem & Proof for the General Case	62
5.2.3	Summary & Insight	67
5.3	Design of Biorthogonal Cosine-Modulated Filter Bank	68
5.3.1	Design Method Based on Theorem	68

5.3.2	Joint-Analysis-and-Synthesis-Window Design	69
5.4	Significance to Audio Compression	69
5.4.1	Filter-Bank Design Criteria in Audio Compression	69
5.4.2	Synthesis-Filter-Bank Design in Audio Coders	71
6	Conclusion	74
6.1	On Biorthogonal Transforms & Filter Banks	74
6.2	On Digital Audio Compression	75

List of Figures

2.1	Masking effect and threshold in quiet.	19
2.2	High-level block diagram of an audio encoder-decoder pair.	21
3.1	System based on a multirate filter bank.	33
3.2	Polyphase representation of a multirate filter bank.	34
3.3	Illustration of time-frequency tiling.	36
3.4	Example illustrating time-frequency tiling with adaptive block size.	37
3.5	Example illustrating time-frequency tiling with nonuniform bandwidths.	38
3.6	Coding of the castanet signal.	39
3.7	Example illustrating arbitrary time-frequency tiling.	40
4.1	General structure of a system based on a lapped transform.	43
5.1	Simplified diagram of a system based on a lapped transform.	63
5.2	Alternative formulation of system (ELT with amplitude equalization).	64
5.3	A different representation of the system in figure 5.1.	66
5.4	Subsystems from figures 5.2 and 5.3.	66
5.5	Analysis and synthesis windows for a special case ($K = 1$) of the biorthogonal cosine-modulated filter bank.	70
5.6	Comparison of synthesis windows used for the orthogonal ELT and the biorthogonal cosine-modulate filter bank.	72

List of Tables

2.1	Critical band rate in relation to linear frequency.	18
5.1	Reduction of constraints for different values of K and s	61
5.2	Conjecture for the special case of $K = 4$ and the general case.	61

Chapter 1

Introduction

We live in a world of ubiquitous sound. Much of our experience is accumulated from information conveyed through acoustic waves. Acoustic events, from the urgent ringing of an alarm clock to the droning sermon of a preacher, when perceived by the human auditory system, are collectively termed “audio signals.” The possibility of reproducing these audio signals, even if they have originated at a distance or in the past, has long fascinated scientists and engineers. Among the pioneering products of their research efforts are Bell’s telephone and Edison’s phonograph. On the heels of these early innovations, an explosive expansion of audio material strained transmission and storage resources. The focus of this dissertation, audio compression, is directed at alleviating this problem.

1.1 Context & Motivation

1.1.1 Rise of Digital Audio

Until recently, processing of audio signals was undertaken exclusively in the analog domain. Upon the advent of digitally recorded sound, new possibilities were introduced. There are many advantages to digital audio. Among them are a reduced dependence on equipment quality, the possibility of having multiple stages of processing without multiple degradations and the availability of the entire gamut of digital processing techniques. In fact, modern audio compression schemes would not have been possible had they not been performed in the digital domain.

The popularization of digital audio should be attributed to the introduction of the commercial Compact Disc (CD) system. The aforementioned advantages of digital audio, coupled with a durable medium, allowed the CD to transplant the analog vinyl record with ease. As a matter of fact, the CD standard—digital audio sampled at 44.1 kHz, and linearly quantized to 16 bits—has become a benchmark for researchers in digital audio compression. The phrase “CD-quality” has come to represent the goal of most audio coder designs.

1.1.2 Need for Audio Compression

The demand for high-quality audio in transmission systems such as Digital Audio Broadcast (DAB) and High-Definition TeleVision (HDTV), as well as commercial products such as the MiniDisc (MD) and the Digital Compact Cassette (DCC), is the impetus behind the considerable interest in audio compression schemes. In these applications, either storage space or channel bandwidth is limited. The objective therefore is to maintain high quality at a rate significantly smaller than the 16 bits/sample used in current CD systems.

1.1.3 Audio Compression & High-Definition Television

Among the major applications of digital audio compression is the transmission of HDTV sound. During the development of HDTV systems, the video aspects, such as image compression, have always been in the limelight. Unfortunately, the attention of researchers and the interested public alike has been so focused on the video component that the second half of a complete television system—the audio component—is often neglected or relegated to the sidelines. Although the portion of the terrestrial broadcast channel occupied by the audio signal can arguably be described as insignificant when compared to that occupied by the video signal, the same cannot be said of the role audio plays in the television experience. It has been observed that television viewers are generally sensitive to relatively small defects in the sound quality even if they are willing to tolerate terribly degraded television pictures. Whether this phenomenon is culturally instilled or biologically inherent, it still shows that the treatment of audio signals deserves attention.

At the Advanced Television and Signal Processing Group, we have specifically studied the application of audio compression schemes to the transmission of HDTV sound. An earlier implementation, the MIT Audio Coder (MITAC), is one of the systems that was considered for inclusion in the United States HDTV standard.

1.1.4 Multidimensional & Multichannel Audio

In the context of HDTV, which calls for spatially realistic sound, multichannel audio becomes an important issue. Although conventional wisdom holds audio signals to be one-dimensional and video signals to be three-dimensional, it is clear that both arises from settings that are four-dimensional—three in space and one in time—in nature. The reproduction of the spatial coordinates of various sound sources in an audio recording substantially enhances its realism. The simplest effort to achieve spatially distinct audio involves the simultaneous recording of two separate and independent

channels—left and right—and their subsequent reproduction. This is in recognition of the human auditory system which has two receptive sensory organs, namely the ears. The dual-channel system has usurped the perhaps overly general label of “stereophonic” sound and is currently standard in commercial and professional audio equipment.

However, in an open speaker-driven setting, such as that of a typical movie theater, the dual-channel system is still inadequate for the reproduction of realistic three-dimensional audio “images.” One obvious solution is the addition of channels to the audio recording. Several systems have been proposed along these lines. Among them, emerging as a possible standard, is the Dolby Surround Sound system which includes the usual left and right channels plus a center dialog channel in the front, two surround channels in the rear and a narrow-band special-effects (subwoofer) channel.

Although there generally exists considerable correlation between channels in a multichannel system and there are also documented psychoacoustic effects in the human perception of stereophonic sound, there are certain advantages to processing the several channels in an independent manner. These include, but are not limited to, prevention of cross talk between channels, and compatibility with simpler systems that have fewer channels. While the issue of multichannel sound is not ignored in this research, the main focus will be on the compression of monophonic sound.

1.2 Dissertation

1.2.1 Scope of Thesis

In this research, we build upon previous efforts in audio research by considering one important aspect of audio coder design, the short-time spectral decomposition, which should give rise to an appropriate and efficient audio signal representation. The contribution of the doctoral dissertation is twofold:

- A survey of the current status of audio compression research is included. Particular emphasis is placed on different spectral decomposition schemes used and the various audio signal representations that arise from their application.
- The focus of this thesis is an investigation of the incorporation of biorthogonality into a cosine-modulated filter bank. Modulated filter banks are popular in audio compression schemes. The particular formulation under study is the extended lapped transform (ELT). A complete mathematical formulation of the new biorthogonal filter bank will be presented. The advantage of incorporating biorthogonality into the ELT is then shown from a theoretical viewpoint,

specifically by establishing the increase in the degrees of design freedom. A suggestion on how this increased flexibility can be used to improve audio compression is also provided.

1.2.2 Organization of Thesis

The thesis will be divided into six chapters which include discussions of the general aspects of audio compression and the more specific details of a biorthogonal transform:

Chapter 1 **Introduction**

Chapter 2 **Digital Audio Compression** This chapter contains a summary of the current state of audio compression research.

Chapter 3 **Multirate Filter Banks & Audio Signal Representations** The different audio signal representations that arise from various spectral decomposition schemes are discussed in this chapter. Since this decomposition is usually achieved by a multirate filter bank, the chapter starts with a brief overview of multirate filter bank theory.

Chapter 4 **Biorthogonality in Lapped Transforms** This chapter contains the mathematical formulation of the biorthogonal cosine-modulated filter bank which is an extension of the extended lapped transform (ELT).

Chapter 5 **Advantages from Incorporation of Biorthogonality** In this chapter, we establish the increase in degrees of freedom from the incorporation of biorthogonality into the ELT. Three special cases ($K = 1, 2, 3$ where K is the overlapping factor) are explored and a general result is shown and proven. The chapter ends with a suggestion of how this increased flexibility can be used in audio compression.

Chapter 6 **Conclusion** The thesis concludes with some final thoughts on the topics of biorthogonal filter banks and digital audio compression.

Digital Audio Compression

The notion of compressing high-quality audio met with much unwarranted initial skepticism, especially from self-styled audio consumer experts, commonly known as “audiophiles.” Some of their fears, however, are understandable and can perhaps be justified. The myriad sources from which audio signals originate preclude a simple source model. This is markedly different from other forms of digital data that have been successfully compressed. For example, speech has an easily accessible production system, the human vocal tract, that can be studied and modeled. Common meaningful images can be expected to have useful spatial correlation, whereas video, with the extra dimension, generally promises temporal in addition to spatial correlation. Unfortunately, general audio signals are not guaranteed to have features that can be similarly exploited. Nevertheless, in recent years, significant progress has been made in audio data reduction by the inclusion of perceptual modeling. While perceptual modeling has also been used in speech and image compression [1], the exceptional emphasis on modeling the receiver rather than the source gives the problem of audio coding a certain uniqueness in the field of compression.

The current state of audio compression research is summarized in this chapter. Given the prominence of psychoacoustic modeling in audio compression schemes, it is discussed in the immediately succeeding section.

2.1 Psychoacoustics

Study of the human auditory system is known collectively as “psychoacoustics.” Years of research have revealed the human ear to be an amazingly versatile and surprisingly adaptive instrument. While the ear can be sensitive to minute details in the acoustic environment, it can also be tolerant of significant levels of distortion in audio signals by rendering them inaudible in the perception process. It is this latter characteristic that is found to be useful in audio compression schemes.

An extensive treatise on psychoacoustics can be found in [2]. For the purpose of this dissertation, a few relevant topics are chosen for more detailed discussion. These include the notion of critical band analysis, the threshold in quiet, and the masking effects.

2.1.1 Critical Band Analysis

The notion of critical bands is central to the study of human auditory perception. The term was first coined by H. Fletcher over 45 years ago [3]. In his experiments, he discovered that for each distinct tone, there exists a band of frequencies centered at that tone in which the just-noticeable noise energy remains nearly constant, regardless of the bandwidth of the noise and its spectral shape. Furthermore, he discovered that the bandwidth of this critical band tends to increase as the frequency of the center tone increases. Roughly speaking, the critical bandwidth remains approximately 100 Hz up to a frequency of 500 Hz. Above that frequency, critical bands show a bandwidth of about 20% of the center frequency. This conforms to the expectation that the frequency resolution of the human ear decreases and its temporal resolution increases with frequency.

For ease of usage, researchers have derived a critical band scale by dividing the linear frequency scale into a bank of adjacent bandpass filters, each with a critical bandwidth. Although actual inner ear operation would not show such distinct demarcations, the resulting scale has been useful for reference purposes. The critical band number, x , is given the unit of bark, after the German scientist Barkhausen. There are around 24 barks from DC to 13.5 kHz. Table 2.1 (taken from [2]) shows the relationship between the bark scale and the linear frequency scale. In [4], a mathematical relationship is given between linear frequency f (Hz) and critical band number x for f above 500 Hz:

$$f = 650 \sinh(x/7). \quad (2.1)$$

In later experiments many different methods were used to derive and verify the critical bandwidths. Each confirmed the validity of the concept. It was also found that the critical band scale is rooted in human anatomy. For example, the critical band rate, x , was found to be approximately proportional to the spatial distance along the basilar membrane; that is, 1 bark \approx 1.3 mm [2]. Physiologically, 1 bark corresponds to roughly 150 hair cells [2] and 1200 primary nerve fibers [4] in the cochlea.

2.1.2 Threshold in Quiet

The threshold in quiet, also known as the absolute hearing threshold, is the lower limit of human hearing. It indicates, as a function of frequency, the intensity of a pure tone that is just audible. In acoustic research, the intensity of sound is measured in terms of sound pressure level (SPL), L .

x	f_l, f_u	f_c	x	Δf_G	x	f_l, f_u	f_c	x	Δf_G
Bark	Hz	Hz	Bark	Hz	Bark	Hz	Hz	Bark	Hz
0	0				12	1720			
		50	0.5	100			1850	12.5	280
1	100				13	2000			
		150	1.5	100			2150	13.5	320
2	200				14	2320			
		250	2.5	100			2500	14.5	380
3	300				15	2700			
		350	3.5	100			2900	15.5	450
4	400				16	3150			
		450	4.5	110			3400	16.5	550
5	510				17	3700			
		570	5.5	120			4000	17.5	700
6	630				18	4400			
		700	6.5	140			4800	18.5	900
7	770				19	5300			
		840	7.5	150			5800	19.5	1100
8	920				20	6400			
		1000	8.5	160			7000	20.5	1300
9	1080				21	7700			
		1170	9.5	190			8500	21.5	1800
10	1270				22	9500			
		1370	10.5	210			10500	22.5	2500
11	1480				23	12000			
		1600	11.5	240			13050	23.5	3500
12	1720				24	15500			
		1850	12.5	280					

Table 2.1: Critical band rate, x , in relation to the lower (f_l) and upper (f_u) frequency limit of critical bandwidths (Δf_G), centered at f_c [2].

The relationship between L and sound intensity I is given by

$$L = 20 \log(I/I_0) \text{ dB} \quad (2.2)$$

where the reference value I_0 is defined as 10^{-12} W/m^2 .

The broken line in figure 2.1 depicts the threshold in quiet for an average person. Over-exposure of the hearing system to loud sounds may cause temporary or permanent shifts in this threshold. As can be seen from the figure, people with normal hearing are most sensitive to soft sounds in the range of 10–18 barks (1–5 kHz). People who are suffering from hearing loss, however, may have an elevated threshold in parts of the spectrum such as around 3–4 kHz.

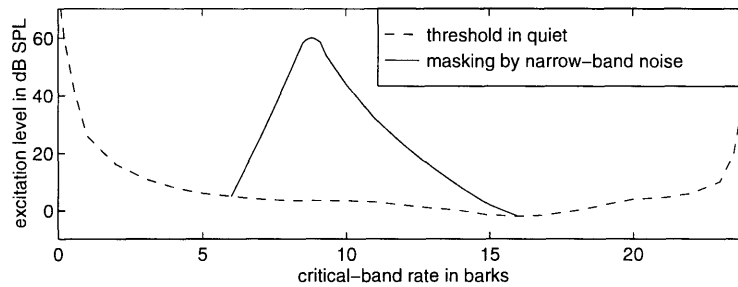


Figure 2.1: Masking effect and threshold in quiet [2]. The broken line is the threshold in quiet. The solid line shows the altered threshold due to narrow-band noise of 60 dB sound pressure level centered at 1 kHz.

2.1.3 Masking Effects

Masking is the phenomenon in which one signal is rendered inaudible by another, typically louder, signal occurring in close spectral or temporal proximity. The signal responsible for the masking effect is termed the *masker* and the signal rendered inaudible is the *maskee*. When the two signals have different spectral contents but occupy the same temporal space, the effect is known as *simultaneous masking* or *spectral masking*. If, on the other hand, the two signals occur at different times, usually in close succession, then the effect is called *non-simultaneous masking* or *temporal masking*.

Exploitation of the masking effects is perhaps the most significant concept in audio compression. The objective is to shape the distortion introduced by quantization and data reduction in such a way that the distortion will be masked and therefore inaudible.

2.1.3.1 Simultaneous (Spectral) Masking

The methods by which the critical bands of section 2.1.1 are derived are deeply rooted in the ideas of spectral masking. Of primary interest though are the masking effects that extend beyond critical bands. When a signal enters the inner ear, its energy is spread along the basilar membrane. Therefore, a softer signal at a frequency close to that of the tone is rendered inaudible. Studies have revealed that different types of masking occur, including tones masked by noise and tones masked by other tones. The phenomenon is complex and masking curves differ significantly in level and shape with the intensity level of the masker. For example, the masking effect due to narrow-band noise centered at 1 kHz is shown in figure 2.1 (adapted from [2]). The presence of the masker noise leads to an elevated just-noticeable threshold for maskee tones that is represented by the solid line in the diagram.

2.1.3.2 Non-Simultaneous (Temporal) Masking

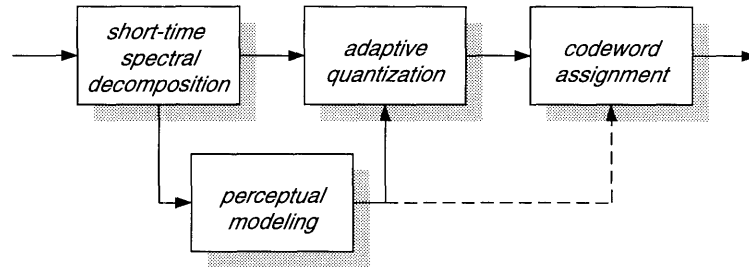
Related but separate from the above are masking effects in the temporal domain. Empirical evidence suggests that a signal can mask weaker signals immediately preceding and succeeding it. There is also evidence that masking for signals occurring after the masker lasts longer than masking for signals occurring before the masker. Furthermore the temporal masking effect differs with frequency. As discussed above, the temporal resolution of the human auditory system increases with frequency. This is reflected in the temporal masking effects which last longer for lower frequency signals than for higher frequency signals.

While it is obvious that knowledge of spectral masking can be applied to the shaping of quantization noise in the spectral domain, some understanding of temporal masking effects is also essential in audio coder design. Because of the time-frequency uncertainty principle, fine frequency analysis can only be achieved at the expense of temporal resolution. By using the extent of temporal masking as a lower limit of temporal resolution, a compromise can be reached so that temporal artifacts are avoided. More on this topic will be discussed in section 3.2.2.

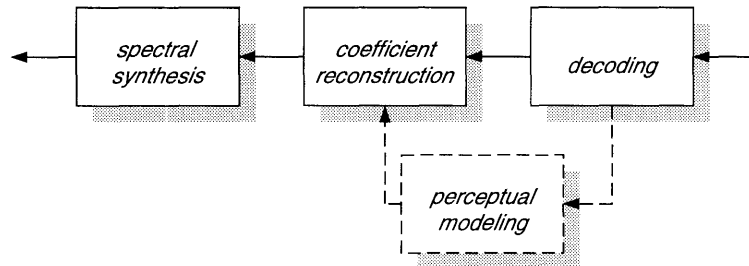
2.2 Frequency-Domain Coding of Audio

Currently, almost all popular audio coders perform compression in the frequency domain. In the signal-processing literature, such schemes generally come under the names, “transform coder” and “subband coder”. Transform/subband coding of audio has much that is similar to the transform/subband coding of other forms of digital data, such as image, video and speech. Figure 2.2 is

a high-level block diagram indicating some of the common elements in an audio encoder-decoder pair. Many of these, such as spectral decomposition and synthesis, quantization and reconstruction, are also found in image and speech coders. In the following sections, we will briefly discuss each of these stages and some of their unique aspects when found in audio compression systems.



(a) encoder



(b) decoder

Figure 2.2: High-level block diagram of an audio encoder-decoder pair. Dotted lines indicate optional elements.

2.2.1 Short-Time Spectral Decomposition

The first stage of frequency-domain coding involves a short-time spectral decomposition of the audio signal. This serves to recast the signal in a domain that is not only amenable to perceptual modeling but also conducive to achieving transform coding gain. The techniques used are traditionally divided into subband filtering and fast transforms, though the underlying concepts are similar. The relatively recent emergence of the field of multirate filter banks presents a unified theory underlying both techniques.

For the purpose of audio compression, several criteria have consistently been regarded as

important in the construction of these decomposition schemes:

1. *Critical Sampling* This means that the aggregate rate of the subband channels, or alternatively the transform coefficients, is the same as the input sample rate. Critical sampling, while not essential, is desirable because it ensures that no extra dependency between samples is introduced due to the transformation into a new representation. Also, subsequent stages of the coder are then not required to operate at a higher aggregate rate than the input sample rate. Critically sampled systems are also called maximally decimated.
2. *Perfect Reconstruction* This refers to signal decompositions from which the original signal can be exactly recovered in the absence of quantization distortion. This again is not essential, but is highly desirable, because it allows the designers to isolate the introduction of signal distortion in the quantization and codeword-assignment modules, thereby simplifying the system design process.
3. *High Frequency Resolution* There are at least two reasons for wanting high frequency resolution. First, when the audio signal has low spectral flatness, a filter bank with high frequency selectivity is required to achieve maximal transform coding gain. Second, successful perceptual modeling is based on critical-band analysis of the audio signal. If the frequency resolution of the decomposition scheme can not resolve the lowest and narrowest critical bands of the human ear, all subsequent computations based on perceptual modeling will be invalid.
4. *High Temporal Resolution* Although it is a conflicting requirement from the above, high temporal resolution is also desirable. Low temporal resolution in the decomposition scheme can lead to artifacts in the compressed signal, the most notable of which is the “pre-echo.” The “pre-echo” issue will be explored in detail in section 3.2.2.

The signal decomposition schemes in most popular audio coders are based on a cosine-modulated filter bank. It is known under several different names in the literature. Among them are “Modulated (or Extended) Lapped Transform (MLT & ELT),” “Time-domain Aliasing Cancellation (TDAC) Filter Bank,” and “Modified Discrete Cosine Transform (MDCT).” Subsequent chapters contain further discussion of cosine-modulated filter banks.

2.2.2 Perceptual Modeling

The objective of perceptual modeling in an audio compression scheme is to compute dynamically a masking threshold under which all distortion is inaudible. Given this masking curve, it is then

possible to mold judiciously the noise that is inevitably introduced by the coding process to achieve significant signal compression with minimal loss of perceived audio quality.

The ideas introduced in section 2.1, especially the masking effects in subsection 2.1.3, are used in the generation of the masking threshold. Earlier, it was mentioned that when a tone enters the inner ear, its excitation power is spread along the basilar membrane surface. Therefore, signals of different frequencies and of lower amplitudes may be perceptually masked. Extrapolating this special case to a general audio signal and coupling it with the absolute hearing threshold (threshold in quiet), it is possible to derive a signal-dependent masking curve which will form the basis for quantization-noise shaping.

It is well established that the quality of audio suffers without perceptual modeling. In [5], the researchers implemented two coders—one with and one without perceptual modeling. Impairments were found to be audible only in the non-perceptual coder. Another audio coder which lacks perceptual modeling was found to deliver good audio quality only at high bit rates [6].

2.2.3 Quantization and Codeword Assignment

Data reduction in audio coders occurs in the quantization stage. After the audio signal has been decomposed, the common practice is to perform dynamic bit allocation. The maximum magnitude of the distortion introduced to each subband channel from quantization is inversely proportional to the number of bits allocated. Therefore, it is possible, given a masking curve, to allocate bits such that the quantization noise is below the perceptual threshold. This bit allocation is typically performed at the encoder and the resulting distribution sent to the decoder as side information. An alternate method is to perform the bit allocation in parallel at the decoder [7]. The advantage of the former method is that, in a broadcast scenario where there are many decoders to one encoder, it is economically more feasible to make adjustments to the bit-allocation algorithm. This, of course, comes at the expense of wasted bandwidth for the transmission of side information, a cost that the latter method can avoid.

There are variations to the popular practice of using uniform scalar quantization on the subband samples. Several past proposals have suggested the use of various forms of vector quantization (VQ) in which the samples are quantized jointly instead of independently [8, 9, 10]. In our research group, experimentation with a simplified form of VQ, known as lattice vector quantization [11], has shown a small but significant coding gain at the expense of added complexity.

The exact digital representation for sending the quantized samples over a transmission channel or for storage in a medium also varies from the simple to the complex. The most straightforward

method is to send directly the bits associated with the quantized sample. A refined method would be to assign to each quantized sample a variable-length codeword using an entropy-coding scheme such as Huffman coding. This, however, necessitates the use of a buffer feedback mechanism to maintain the channel rate [11].

2.3 Other Audio Coding Schemes

The above discussion has provided a brief overview of a conventional audio coder in which compression is performed in the frequency domain. Not surprisingly, there are other audio compression structures. While it is safe to say that the frequency-domain coding structure described here is by far the most popular and probably the most effective, it is worth noting some other coding schemes that differ from the above generic form.

2.3.1 Linear Predictive Coders

Linear Predictive Coding (LPC) is a popular speech compression technique. Using this method, considerable source-coding gain can be realized by accurately predicting the speech samples based on simple source statistics. However, typical audio signals, as we have noted earlier, do not usually follow any particular source model. Therefore, LPC performs poorly when applied to audio compression. In [12], it was demonstrated that the high degree of non-stationarity of the audio signals leads to a loss of prediction gain.

It is worth noting that a variation of LPC known as multi-pulse LPC has been used to code wideband audio and the researchers who proposed the improved method have claimed a certain amount of success [13, 14, 15].

2.3.2 Subband ADPCM Coders

In the technique known as subband ADPCM (Adaptive Differential Pulse Code Modulation), the audio signal is first decomposed to a low number of subbands, typically less than four, and then encoded by adaptively predicting the samples. The CCITT G.722 coding standard is the most prominent example of this compression method. It uses no perceptual modeling for noise shaping and therefore is mainly suitable for coding wideband speech with a bandwidth of up to 7 kHz (sampled at 16 kHz) at bit rates of 64, 56 and 48 kbits/s.

Other researchers have attempted to improve upon the G.722 standard by increasing the number of subbands and by using some coarse perceptual modeling. Examples of these efforts are found in [16] and [17].

2.3.3 Dynamic Dictionary

Another interesting audio compression scheme was proposed in [18, 19]. In this approach, the encoder and the decoder maintain a dynamic dictionary of audio waveforms. Only the difference between the actual signal and the corresponding dictionary entry is transmitted using a wavelet-based transform coding scheme. This is, in some respects, similar to predictive coding. One significant difference comes from the use of time warping for matching the waveform to the dictionary entries. This allows the coder to take differences in time-scale into account. The performance of this coder is allegedly excellent at 48–64 kbits/s.

2.4 Multichannel Audio Compression

Although the study of multichannel audio is beyond the scope of this thesis, we would still like to point out briefly two of the coding techniques that are used to compress stereo and surround-sound material.

2.4.1 Sum-Difference Stereo & Matrix Surround

The idea here is simple. When the method applied to a dual-channel audio signal, coding is performed on the derived channels, $L + R$ and $L - R$ where L and R represent the original left and right channels respectively. This rematrixing of the audio channels is effective in removing some of the correlation between the left and right channels. The idea of rematrixing before compression was first used in [20]. It was extended to surround sound in [21, 22]. A danger of coding in the new domain is that the original perceptual models may no longer be applicable.

2.4.2 Intensity Stereo & Intensity Surround

Intensity stereo is a technique that exploits another aspect of the human auditory system. In the perception of stereophonic material, the human ear is found to lack the ability to detect phase differences between channels at high frequencies. Therefore, it is only necessary to send the envelope

information of the audio signal and then scale it accordingly for each channel. This compression method was used in the MPEG standards (see following section) for coding stereo and surround-sound material.

2.5 Major Research & Development Efforts in Audio Compression

2.5.1 Erlangen, Germany

In Germany, centered at the University of Erlangen is a large research effort housing many of the pioneers of high-quality audio compression. Prior to their participation in the effort to develop ASPEC and the MPEG standards, the researchers there, under Karlheinz Brandenburg, developed two frequency-domain coding algorithms: Low-Complexity Adaptive Transform Coding (LC-ATC) [23, 24] and Optimum Coding in the Frequency domain (OCF) [25, 26, 27]. Aside from algorithmic design, the group is also involved in real-time implementation issues [23, 24, 28, 29].

2.5.1.1 Low-Complexity Adaptive Transform Coding (LC-ATC)

Classical Adaptive Transform Coding (ATC) [30] was developed for the compression of speech. Modification to LC-ATC involves an inclusion of perceptual modeling and also a simplification. The bit rate required is slightly higher than previous ATC designs. In the LC-ATC compression scheme, either a block transform or a lapped transform is used to derive 512 spectral coefficients which are then divided into 46 groups according to the critical bandwidths. A portion of the available bits follow a fixed allocation while the rest are dynamically allocated using a perceptual model derived from a logarithmically quantized spectral envelope transmitted as side information. Quantization of the spectral coefficients is by a block companding method (block floating point).

2.5.1.2 Optimum Coding in the Frequency Domain (OCF)

The OCF algorithm underwent several improvements since its inception. The latest incarnation uses a 1024-point lapped transform to derive 512 spectral coefficients. The perceptual modeling stage takes into account the threshold in quiet, basilar-membrane spreading and differences between tonal and noise-like signals. For coding of the coefficients, a non-uniform quantizer is used in conjunction with variable-rate entropy coding. The quantization and the codeword assignment are done in two iteration loops. The outer loop uses an analysis-by-synthesis method to constrain the quantization noise to below the dynamic masking curve by altering the step sizes in individual

critical-band groups. The inner loop adjusts the overall step size to maintain a constant bit rate. The researchers also took note of the temporal artifact known as “pre-echo” (see section 3.2.2 for details). A variety of control measures were tried. One interesting proposal performs pre-filtering on the audio frame which contains the transient signal.

2.5.2 AT&T Bell Laboratories

In the United States, one of the major thrusts in audio compression research comes from AT&T Bell Laboratories. James Johnston was responsible for the Entropy-coded Perceptual Transform Coder (PXFM) [31]. The PXFM coder uses a 2048-point DFT with 1/16 overlap between blocks for analysis. Its major contribution though is the development of a perceptual model which is also used in the OCF. Johnston detailed the method for generation of a masking curve in [31]. His procedure takes into account critical band analysis, tonality estimation, basilar membrane spreading and absolute hearing thresholds. With this model, Johnston was able to derive an entropy estimate for perceptually transparent audio compression, which he termed “perceptual entropy” [32]. Johnston later extended the PXFM to perform sum-difference encoding on wideband stereo signals. This is known as the Stereo Entropy-coded Perceptual Transform Coder (SEPFM) [20].

Currently, AT&T is advancing the next generation of Perceptual Audio Coder (PAC) which incorporates surround-sound compression and variable time/frequency resolution. This scheme will be used in at least one of the entrants in the the DAB-standard competition [33] in the United States.

2.5.3 Centre National d’Etudes des Télécommunications (CNET)

In France, the Centre National d’Etudes des Télécommunications (CNET) also studied a form of adaptive transform coder which is similar to the ones described above [34, 35, 36]. There are two interesting variations. In the quantization and codeword-assignment stages of their compression scheme, the coefficients that have energies below the masking threshold are simply not coded. Instead, their location in the spectrum, or indices, are transmitted using run-length encoding. The other variation concerns the transmission of the spectral envelope, or spectrum descriptor, which is necessary as side information to reproduce the masking threshold at the decoder. A predictive scheme that exploits the correlation between successive temporal blocks is used.

In addition to transform coding, the center also proposed a subband ADPCM coder [16] at an earlier date.

2.5.4 Adaptive Spectral Entropy Coding (ASPEC)

Adaptive Spectral Entropy Coding (ASPEC) is the joint effort of several parties, including the research groups at Erlangen, AT&T Bell Laboratories and CNET [37]. It is supposed to combine the best parts of OCF, PXF, and several other transform coders. The basic scheme uses a 512-band MDCT for analysis and two levels of psychoacoustic modeling. The spectral envelope is sent as side information using the predictive scheme proposed by CNET. Pre-echo control is by means of adaptive block-size switching. Quantization and codeword assignment are accomplished using a double-loop scheme similar to that of OCF.

ASPEC is conceived as an alternative to MUSICAM when the MPEG standards were defined. A collaborative effort merging the two systems later resulted in layer III of the audio portion of the MPEG standard.

2.5.5 Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing (MUSICAM)

Masking pattern adapted Universal Subband Integrated Coding And Multiplexing (MUSICAM) was developed by the Centre Commun d'Etudes de Télédiffusion & Télécommunications (CCETT), the Institut Für Rundfunktechnik (IRT), Matsushita, and Philips under the European project known as Eureka 147 [38]. The basic structure of the coder is simple. Subband decomposition is used to derive 23 subband channels. The masking threshold is calculated in parallel using a 1024-point FFT. For quantization, bits are dynamically allocated to blocks of 36 samples in each subband and three scale-factors are transmitted for each block. The coding scheme is divided into three upwardly compatible layers with increasing complexity and increasing subjective performance.

2.5.6 Moving Pictures Expert Group (MPEG)

The Moving Pictures Expert Group is a committee operating within the International Organization of Standardization (ISO/MPEG). They are responsible for developing a series of audio-visual standards for the electronic computing environment, high-definition television, and teleconferencing. Its audio coding standard is the first international standard in the field of high-quality digital audio compression [39, 40].

The standardization process is heavily influenced by two existing compression schemes, MUSICAM and ASPEC. For example, the idea of having three layers is an inspiration by MUSICAM. In fact, layers I and II of MPEG are basically those of MUSICAM. In layer III, though, there is

an effort to increase the frequency resolution of the analysis filter bank. A hybrid filter bank that cascades MDCT's upon the 23 original subbands is used to maintain compatibility. Other features such as entropy coding and run-length encoding for zero-value coefficients are also incorporated. There are also modes for separate treatment of stereo and surround-sound material.

In Phase II of the MPEG standardization, other non-backwards-compatible compression schemes will probably be incorporated into the standard. These might include the Dolby AC family of coders and PAC from AT&T.

2.5.7 Digital Compact Cassette (DCC) & MiniDisc (MD)

At the time of writing, two consumer electronic products on the market use audio compression schemes: the Digital Compact Cassette (DCC) from Philips and the MiniDisc (MD) from Sony. Philips has previously been very active in audio research. Earlier efforts include the development of MUSICAM, the study of subband audio coding [41] and the compression of stereophonic material [42]. For the DCC, a simplified version of layer II of the MPEG scheme [43] is used. On the other hand, the MiniDisc System from Sony uses a different compression scheme called the Adaptive Transform Acoustic Coder (ATRAC) [44]. This coder uses a hybrid filterbank which first divides the audio signal to three nonuniform subbands. This is followed by dynamically windowed MDCT's. Quantization is by floating-point companding. Both the DCC and the MD have low compression ratios at around 4:1.

2.5.8 Dolby Laboratories

Dolby Laboratories is another major force in the world of high-quality audio. Their noise reduction system and their surround-sound configuration have both become *de facto* standards [45]. For audio compression, researchers at Dolby have developed a family of adaptive transform coders, the AC family, that are suitable for audio-visual applications [46, 7]. The AC-2 and AC-2A audio coders from Dolby use an evenly stacked TDAC filterbank. The AC-2A coder uses window switching for pre-echo control [47]. For the AC-2 coders, quantization and codeword assignment are done on critical-band groupings of spectral coefficients. In AC-3, the originally monophonic coder was extended to manage 5.1 channels of surround sound [48]. The quantization and codeword assignment stages were also modified to allow variable time-frequency groupings of coefficients. AC-3 has been chosen as the audio standard for the Grand Alliance HDTV system which itself will soon be ratified as the television transmission standard of the United States.

2.5.9 MIT-Advanced Television and Signal Processing Group (ATSP)

Research in audio compression here at MIT's Advanced Television and Signal Processing (ATSP) group has been geared towards HDTV applications. An earlier effort, the MIT audio coder (MITAC), was one of the systems competing to become the US HDTV standard and subsequently was considered for inclusion into the Grand Alliance HDTV system [49]. MITAC is also an adaptive transform coder which uses a 1024-point oddly stacked TDAC filter bank. Its unique features include a finely quantized spectral envelope which is differentially entropy encoded.

Subsequent to MITAC, another compression scheme using hierarchical nonuniform filter banks and lattice vector quantization was studied [11]. This is described in more detail in section 3.2.2. The work contained in this dissertation is also part of the research effort of this group.

2.6 Parameters for Assessing Audio Coders

Given the formidable array of different audio compression schemes in the previous section, which still by no means represent an exhaustive list, competition between the various research groups is inevitable, especially when standardization is involved. The correct way to evaluate the quality of an audio coder is therefore a subject of much debate. While it is difficult to agree upon one approach, any good assessment should take into consideration several aspects of the audio coder. A few of these are discussed in the following.

2.6.1 Subjective Quality

Subjective quality of an audio coder is difficult to describe unless the compressed signal is placed against a reference high-quality signal. At the time of writing, 16-bit linearly quantized audio at 44.1 kHz (Compact Disc Quality) or 48 kHz (Digital Audio Tape Quality) is the reference against which all audio quality is measured.

One widely used subjective measure is known as the Mean Opinion Score (MOS). In this test, subject listeners are asked to classify the quality of coders on an N -point scale, the most popular one in use being a 5-point adjectival scale. For the evaluation of each segment of audio, the subject is given three signals (triple stimulus), the first one of which is always the reference. The next two signals are the reference and the compressed signal in a random order (hidden reference) known to neither the conductor of the test nor the listener (double blind). The subject is then asked to decide which of the last two signals is compressed and to give their quality assessment.

Although the MOS values depend very much on the audio material chosen for the test and are difficult to duplicate at different locations, it is found to be fairly reliable for comparison purposes. Given experienced listeners, it is also sensitive to minor impairments in the compressed audio.

2.6.2 Objective Measurements

Degradations to compressed digital audio, being very different from impairments to analog sound, have rendered most of the traditional objective measurements such as total harmonic distortion (THD) and signal-to-noise ratio (SNR) ineffective. Among the newly developed objective measurements is the noise-to-mask ratio (NMR) [50, 51]. The NMR measures the difference between the masking threshold derived from the input signal and the actual distortion introduced by the compression scheme. Another form of evaluation known as the Perceptual Audio Quality Measure (PAQM) [52] transforms the input and output signals of an audio coder to a psychophysical domain before comparison. Both NMR and PAQM have shown high correlation with MOS assessments.

2.6.3 Bit Rate, System Delay & Complexity

Other parameters for assessing the quality of an audio coder are common to all signal compression schemes. Most important among these is the bit rate. Most high-quality audio coders claim to achieve “transparent” quality at a bit rate below 128 kbits/sec. For multichannel sound, there are coders, such as Dolby AC-3, that can achieve a bit rate of under 400 kbits for 5.1 channels. System delay and complexity can also be important. The former is especially important for real-time broadcasting applications. Implementation complexity, though a largely economical issue, also cannot be ignored.

Multirate Filter Banks & Audio Signal Representations

Historically, signal compression schemes that depend on an initial short-time spectral decomposition are classified into transform coding and subband coding. While the two formulations approach the problem from different angles, they are merely different aspects of the same solution. In both cases, the new representation of the signal after decomposition is no longer purely temporal in nature. Transformation to a representation that reflects the frequency components of the signal allows the easy removal of statistical redundancy for the purpose of compression. This is known as transform coding gain. Aside from transform coding gain, the spectral analysis is also essential in audio compression schemes for the exploitation of psychoacoustic results. For example, quantization-noise shaping based on perceptual modeling is performed in the spectral domain.

Not surprisingly, among the most important issues in the design of an audio coder is the selection of a decomposition scheme that will lead to an appropriate and efficient audio signal representation. As mentioned earlier, the signal decomposition is commonly achieved by a multirate filter bank or, equivalently, a lapped transform. In recent years, much attention has been devoted to the study of these filter banks [53, 54, 55]. The progress in the field of multirate filter banks has led to parallel formulations in audio compression schemes. We will therefore devote this chapter to a brief description of these filter banks.

3.1 Theory of Multirate Filter Banks

3.1.1 Basics

The study of multirate filter banks is best approached from the angle of subband filtering. Figure 3.1 shows a generic system based on an M -band multirate filter bank. The analysis stage is composed of M filters, $H_0(z)$ to $H_{M-1}(z)$, while the synthesis stage consists of another set of M filters, $F_0(z)$ to $F_{M-1}(z)$. On the analysis side, decimation by a factor of M is necessary to maintain critical sampling, a criterion first mentioned in section 2.2.1. In this particular example, the bandwidths of

the resultant subband signals are assumed to be uniform and close to π/M . However, it is easy to conceive of nonuniform filter banks, in which case the decimators for each channel will be different. For synthesis, the filtered subband signals are upsampled to the original rate and summed. Under the requirement of perfect reconstruction, also first mentioned in section 2.2.1, the recovered signal $y[n]$ should be, aside from a possible delay, identical to the original signal $x[n]$.

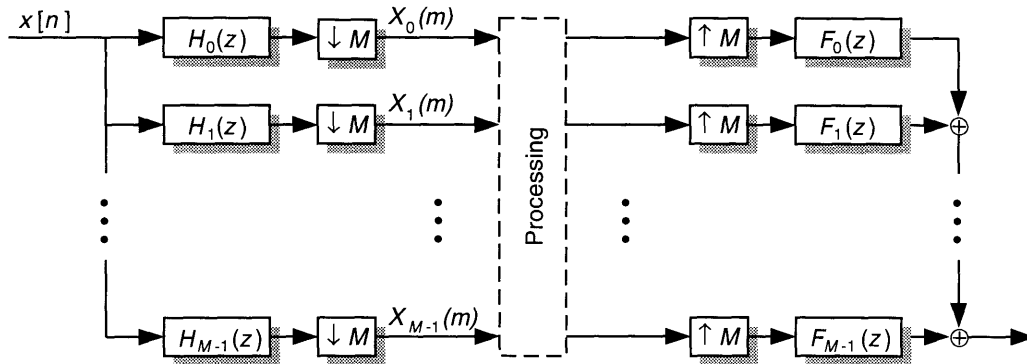


Figure 3.1: System based on a multirate filter bank.

3.1.2 Polyphase Representation

By rearranging the filter banks in figure 3.1 into their polyphase components, a more efficient and insightful representation of the same system emerges. The polyphase representation is shown in figure 3.2. Note that computational efficiency has increased because the decimation is done before the filtering while the upsampling is done afterwards. The polyphase component matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$ are related to the filters $H_i(z)$ and $F_i(z)$ in the following manner.

$$H_k(z) = \sum_{r=0}^{M-1} z^{-r} E_{kr}(z^M) \quad (3.1)$$

$$F_k(z) = \sum_{r=0}^{M-1} z^{-r} R_{kr}(z^M) \quad (3.2)$$

where the subscript kr means the (k, r) th element of the transfer-function matrices.

As can be expected, there is a large collection of $H_i(z)$ and $F_i(z)$ that will satisfy perfect reconstruction. However, the polyphase structure suggests one approach by which an FIR filter bank can be constructed. In [56] and [57], Vaidyanathan selects invertible transfer-function matrices

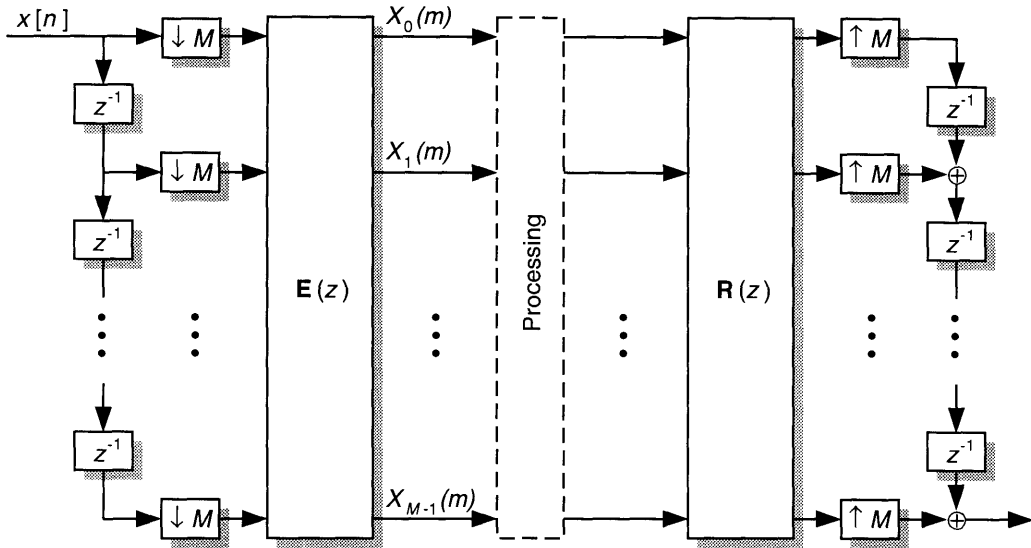


Figure 3.2: Polyphase representation of a multirate filter bank.

that satisfy the following *paraunitary* or *lossless* property to be $\mathbf{E}(z)$.

$$\mathbf{E}(z)\tilde{\mathbf{E}}(z) = \tilde{\mathbf{E}}(z)\mathbf{E}(z) = \mathbf{I} \quad (3.3)$$

where

$$\tilde{\mathbf{E}}(z) = \mathbf{E}^T(z^{-1}).$$

If the synthesis polyphase matrix is then set according to the following, perfect reconstruction is assured.

$$\mathbf{R}(z) = z^{-(N-1)}\tilde{\mathbf{E}}(z) \quad (3.4)$$

where $N - 1$ is the order of the matrix $\mathbf{E}(z)$.

3.1.3 Block Transform

In the context of compression, one can view transform coding—in particular when block transforms are used—as a subset of subband coding in which the filter lengths are equal to the decimation factor. In this scenario, the polyphase component $\mathbf{E}(z)$ is a zeroth-order square matrix and can be

given by

$$\mathbf{E}(z) = \mathbf{A}^T \text{ and } \tilde{\mathbf{E}}(z) = \mathbf{A}. \quad (3.5)$$

The lossless property then implies that the matrix is orthogonal and represents a reversible finite transform. Again there are infinitely many invertible matrices \mathbf{A} that can serve in this context, but in practice only a few structured matrices are used. These include the familiar discrete Fourier transform (DFT) matrix and the discrete cosine transform (DCT) matrix. For example, the DCT-IV matrix is given by

$$a_{ij} = \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right] \quad (3.6)$$

where a_{ij} denotes the (i, j) th element of \mathbf{A} .

Under the block-transform concept, the signal is viewed as being segmented into nonoverlapping blocks and then transformed. This is in contrast to the subband filtering view, in which the signal is filtered and then decimated. The two operations are, however, equivalent in many aspects. While block transforms only form a subset of subband coding, the notion of lapped transforms which will be discussed in the next chapter is much more general. As a matter of fact, in its most general form, the lapped transform is equivalent to a uniform paraunitary FIR filter bank.

3.1.4 Time-Frequency Tiling & Complete Representations

One can also view a multirate filter bank as an implementation of a linear series expansion of a discrete-time signal using an infinite basis, the individual elements of which are localized in time and in frequency [58]. If the basis is complete then the representation will be critically sampled and perfect reconstruction can be achieved. For example, in this view, time-shifted versions of the columns of the transform matrix \mathbf{A} will form the basis functions for the decomposition by that particular block transform.

Localization of the basis functions in time and frequency leads directly to the concept of time-frequency tiling. It is instructive to think of each basis function as having a certain effective region of support in the time-frequency plane—the idea of a time-frequency “tile.” For example, a decomposition scheme with uniform bandwidths essentially transforms a purely temporal tiling of the time-frequency plane into a regular tiling pattern shown in figure 3.3. As we shall see, these diagrams are helpful for visualizing the different decomposition schemes in various audio coders.

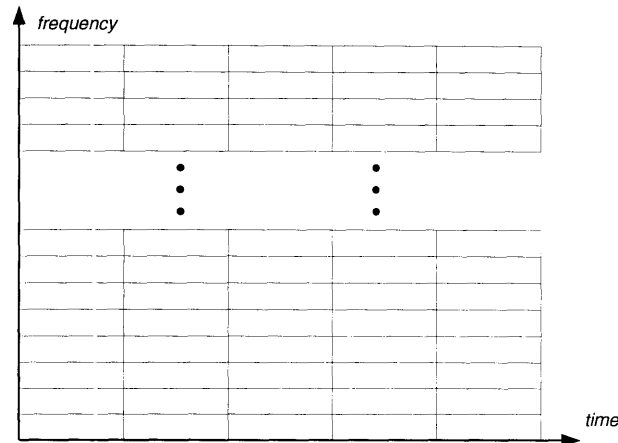


Figure 3.3: Illustration of time-frequency tiling. The particular example shown is a conventional decomposition scheme with uniform bandwidths.

3.2 Multirate Filter Banks in Audio Compression

3.2.1 From Disjoint to Overlapping Temporal Frames

In the context of audio compression, it is, as mentioned earlier, highly desirable for the filter banks to satisfy the twin criteria of critical sampling and perfect reconstruction. The simplest such formulation would be a block transform, such as a DFT filter bank with disjoint temporal frames, which is equivalent to a maximally decimated short-time Fourier transform. The major drawback for this type of decomposition is the occurrence of blocking artifacts which manifest themselves as discontinuities across the edge of temporal frames after quantization. In some early audio coders, critical sampling is compromised to introduce a small amount of overlap between adjacent frames [31]. Other coders, typically those based on subband filtering, sacrifice perfect reconstruction to achieve basically the same result.

The development of lapped transforms [59] and discovery of other critically-sampled perfect-reconstruction filter banks [60] opened new horizons in audio compression. In these decomposition schemes, overlapping between temporal frames is achieved without compromising either critical sampling or perfect reconstruction. Such schemes not only minimize the possibility of blocking artifacts, they also allow the design of better analysis filters, directly affecting the performance of the audio coder.

3.2.2 From Uniform to Nonuniform Filter Banks

Until recently, the representations used by many audio coders, for example Dolby AC-2, ASPEC and MUSICAM [7, 37, 61], have uniform analysis bandwidths. The time-frequency tiling shown in figure 3.3 is a perfect illustration of this. We can see that the simple tiling results in a tradeoff between time and frequency resolution that does not respect the critical bandwidths of the human auditory system. Typically the analysis bandwidth is chosen to be narrow enough to resolve approximately the critical bands in the low frequencies. In the higher frequencies, subbands are grouped to imitate the wider critical bands for perceptual-modeling purposes.

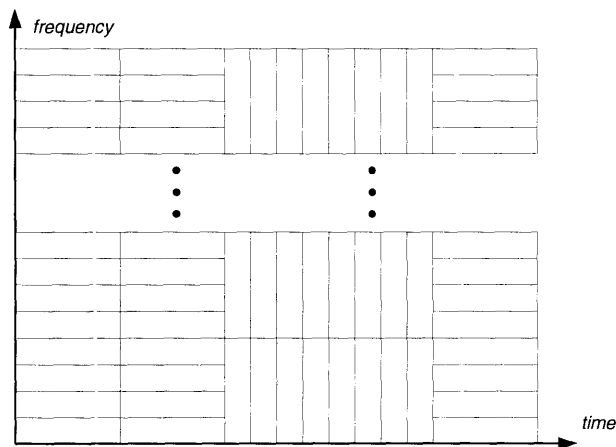


Figure 3.4: Example illustrating time-frequency tiling with adaptive block size.

Unfortunately, the narrow analysis bandwidths lead to poor temporal resolution. One immediate consequence is the occurrence of the temporal artifact known as the “pre-echo.” This particular distortion arises when a sharp transient occurs in the input signal, for example, during a glockenspiel strike in which a sudden burst of energy appears in an otherwise quiet passage. After the coding process, quantization results in noise that is spread over the length of the basis functions. Given that this noise is equally present in both the part of the signal which has high energy and the quiescent part preceding it, the distortion in the quiescent part will be relatively large. If the distortion extends beyond the temporal masking provided by the attack, a “pre-echo” is formed. In some cases, the audible effect is an unpleasant hiss immediately preceding a sharp attack.

There have been various efforts to correct this deficiency. For example, some coders include an adaptive mechanism for improving temporal resolution by adjusting the duration of the temporal frame when a transient is detected [47] (See figure 3.4). This method is favored in many practical implementations because of its simplicity and effectiveness. However, it requires an explicit mech-

anism for detecting transient signals and the compression scheme has to suffer a loss of frequency resolution for the shorter transform blocks.

Another more fundamental approach is the use of a nonuniform filter bank which allows a closer approximation to the critical bands of the human ear (See figure 3.5). Such an approach will allow a more uniform architecture without an explicit adaptation step.

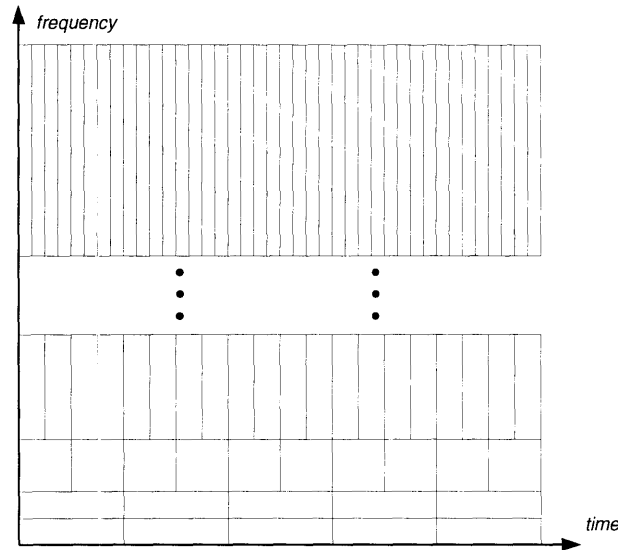
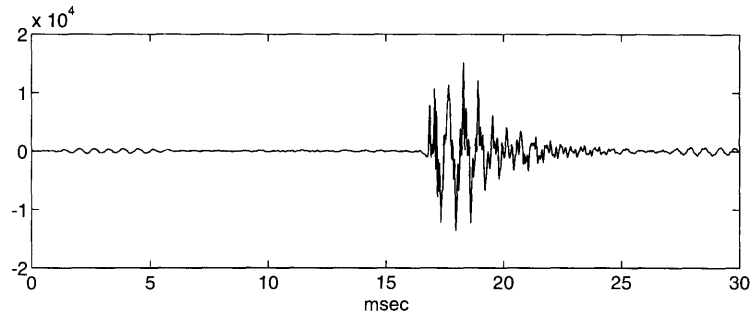


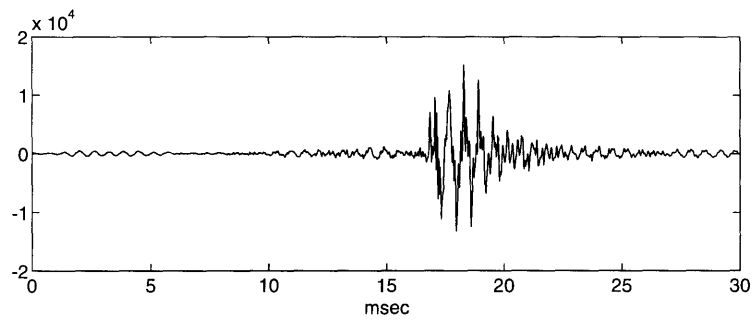
Figure 3.5: Example illustrating time-frequency tiling with nonuniform bandwidths.

Development along these lines includes the use of wavelet representations [18] and fractional-band wavelets to mitigate the coarse octave-scale resolution of two-band wavelets [62]. In this research group, we have experimented with a hierarchical filter bank structure [11] based on the extended lapped transform (ELT), the previously mentioned M -band perfect-reconstruction cosine-modulated filter bank developed by Malvar [63]. For decomposition of audio signals, the hierarchical filter bank offers significant advantages. The fact that various values of M can be used at different levels of the decomposition tree provides flexibility in design and simplifies the task of matching the filter banks to the critical-band structure of the human ear. Similar architectures were developed by Aware Inc. [64] and Malvar [65].

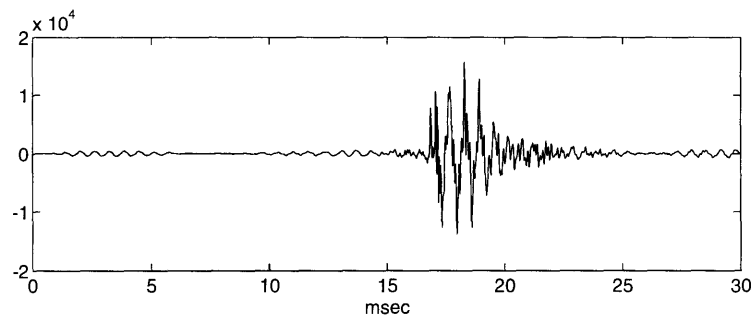
The performance of the hierarchical filter bank for transient signals can be seen in figure 3.6. In this example, the audio signal from a castanet is compressed by two schemes, one using a uniform filter bank and the other using a nonuniform filter bank. In order to achieve a fair comparison, the quantization and coding modules of both systems are adjusted to the same rate. We can observe that the high-frequency components in the pre-echo are attenuated in the signal coded using a



(a)



(b)



(c)

Figure 3.6: Coding of the castanet signal. (a) original signal, (b) version coded with a uniform filter bank, and (c) version coded with a hierarchical nonuniform filter bank.

nonuniform filter bank. An informal subjective listening session shows that the “pre-echo” artifact is inaudible when the signal is coded with the nonuniform hierarchical filter bank.

3.2.3 The Emergence of Time-Varying Filter Banks

The latest advance in filter bank theory stems from time variation of the filter bank. When applied to audio compression, time variation allows the filter bank to adapt to the input signal, thereby achieving a higher compression ratio. There are different aspects in a filter bank that can be time-varying. Previous developments have already considered various possibilities. For example, in [66], the duration of the transform frames varies with the audio signal. This effectively changes the resulting number of subbands from time frame to time frame. In another scheme [19], the basis functions are made to vary while the decomposition structure and therefore the number of subbands remain unmodified.

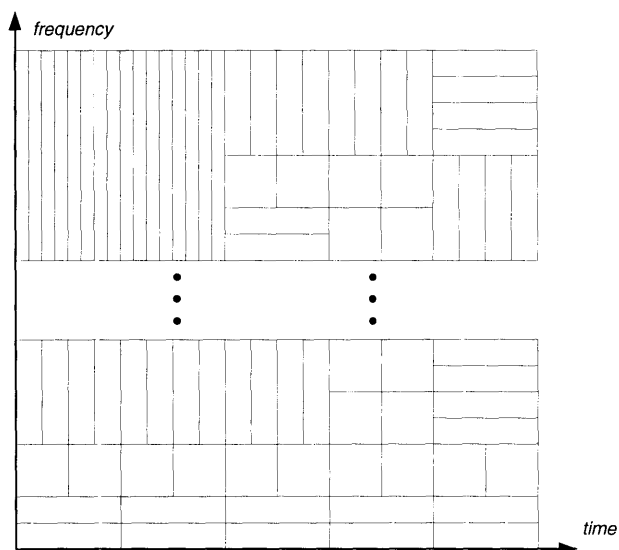


Figure 3.7: Example illustrating arbitrary time-frequency tiling.

Sometimes it is also useful to vary the decomposition structure itself. Consider the audio signal representation in figure 3.5. While that particular filter bank offers a close match to the human auditory system and is therefore mostly free of perceivable artifacts, the wide subbands in the higher frequencies prevent the system from fully realizing the transform gain. A better trade-off between fidelity to the auditory model and maximization of transform gain can be achieved through an adaptive tiling of the time-frequency plane. This concept is best described in [67] and [68] where binary decomposition trees are pruned according to a rate-distortion measure. Figure 3.7

shows a conceptual picture of an arbitrarily tiled time-frequency plane. Given the work done in time-varying lapped transforms [69, 70], extension of the adaptive time-frequency tiling from binary decomposition trees to one with arbitrary numbers of branches at each node should also be possible. Further work on time-varying filter banks can be found in [71, 72, 73, 74]. In addition, audio coders based on adaptive time-frequency tiling recently appeared in the literature [75].

Biorthogonality in Lapped Transforms

In the previous chapter, we have briefly but repeatedly mentioned the lapped transform, a popular family of critically sampled perfect-reconstruction filter banks developed by Malvar. Although lapped transforms are uniform filter banks, they can be fairly versatile. For example, in section 3.2.2, the transforms are cascaded in a hierarchical structure to yield a composite filter bank with nonuniform subbands. In section 3.2.3, the possibility of using time-varying lapped transforms for arbitrary time-frequency tiling is also mentioned.

For this doctoral research, one particular realization of lapped transforms is singled out for improvement by the incorporation of biorthogonality. Previous work in this direction is described in [77, 78, 79] in which the lapped orthogonal transform (LOT) [59] is generalized to become the biorthonormal lapped transform (BOLT). We, on the other hand, are primarily interested in improving the extended lapped transform (ELT) [63]. Part of this work was reported in [80].

4.1 Lapped Transforms

Figure 4.1 shows the general structure of a system based on a lapped transform. The implementation is that of an M -channel filter bank with an overlapping factor of K . The duration of each transform frame or, equivalently, the length of each analysis filter is $2KM$ samples. Note that the decimation factor for each channel, similar to the multirate filter banks discussed in the previous chapter, is M , which is equal to the total number of channels, thereby yielding a critically sampled filter bank.

In figure 4.1, the forward transform is represented by a $2KM \times M$ matrix \mathbf{P} and the inverse transform is represented by a similar matrix \mathbf{Q} . For ease of later development, we introduce \mathbf{P}_l which denotes the l th $M \times M$ square block of \mathbf{P} ; that is,

$$\mathbf{P}^T \equiv [\mathbf{P}_0^T \mathbf{P}_1^T \cdots \mathbf{P}_{2K-1}^T].$$

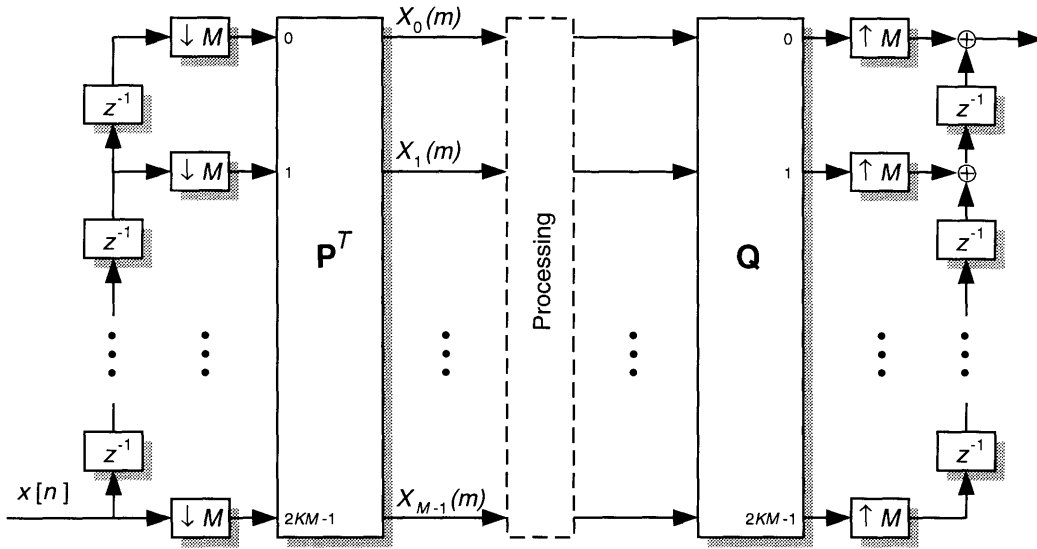


Figure 4.1: General structure of a system based on a lapped transform.

Similarly, \mathbf{Q}_l denotes the l th M by M square block of \mathbf{Q} ; that is,

$$\mathbf{Q}^T \equiv [\mathbf{Q}_0^T \mathbf{Q}_1^T \cdots \mathbf{Q}_{2K-1}^T].$$

In this notation, it is easy to see that the general lapped transform is actually equivalent to a uniform paraunitary FIR filter bank. Recall from section 3.1.2 that the polyphase representation of a multirate FIR filter bank is given by the transfer-function matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$. The relationship between the lapped-transform matrices and the polyphase-component matrices is given by

$$\mathbf{E}(z) = \sum_{l=0}^{N-1} z^{-(N-1-l)} \mathbf{P}_l^T \mathbf{J} \quad (4.1)$$

$$\mathbf{R}(z) = \sum_{l=0}^{N-1} z^{-(N-1-l)} \mathbf{Q}_l^T \mathbf{J} \quad (4.2)$$

where \mathbf{J} is the counter-identity matrix, given by

$$\mathbf{J} \equiv \begin{pmatrix} 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 1 & 0 & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}. \quad (4.3)$$

4.1.1 Extended Lapped Transforms & Cosine-Modulated Filter Banks

For this research, however, we are mainly interested in the particular realization of the lapped transform that implements a cosine-modulated filter bank. Imposing a cosine-modulation structure vastly reduces the design and implementation complexity of the filter bank [81]. We shall see that only one prototype filter, also known as the analysis or the synthesis window, needs to be designed. More importantly, there are fast implementation algorithms available [82, 83, 84]. This type of filter bank is therefore widely used in audio compression. As mentioned before, they are also known as the “time-domain aliasing cancellation” (TDAC) filter bank [85] or the “modified discrete cosine transform” (MDCT) [37]. Both are variations on the same theme. For his realization of a cosine-modulated filter bank, Malvar has coined the terms, extended lapped transform (ELT) and modulated lapped transform (MLT); the MLT corresponds to a special case ($K = 1$) of the ELT.

Using the same matrix notation as above, the forward ELT is given by

$$p_{nk} = h[n] \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right] \quad (4.4)$$

where p_{nk} is the (n, k) th element of \mathbf{P} , and the inverse transform is similarly given by

$$q_{nk} = f[n] \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right]. \quad (4.5)$$

In the above notation, it is conventional to refer to the prototype filters, $h[n]$ and $f[n]$, as the analysis window and the synthesis window, respectively.

4.1.2 Perfect Reconstruction in the Extended Lapped Transform

The ELT originally conceived by Malvar is an orthogonal transform in which

$$\mathbf{P} = \mathbf{Q} \quad \iff \quad h[n] = f[n]. \quad (4.6)$$

Aside from equating the analysis window and the synthesis window, Malvar further stipulates that the windows used are symmetric, such that

$$h[n] = h[2KM - n - 1], \quad f[n] = f[2KM - n - 1]. \quad (4.7)$$

Under these assumptions, time-domain analysis [86] shows that perfect reconstruction is achieved if both the analysis and synthesis windows satisfy the following constraint:

$$\sum_{m=0}^{2K-1-2s} h[mM + n]h[(m + 2s)M + n] = \delta(s) \quad (4.8)$$

for $s = 0, \dots, K - 1$ and $n = 0, \dots, \frac{M}{2} - 1$. The set of nonlinear equations represents $KM/2$ independent conditions on the KM different coefficients in the analysis window. This leaves $KM/2$ degrees of freedom for design purposes.

4.2 Generalization of the Extended Lapped Transform

Relaxation of the requirement in (4.6)—the analysis window equals the synthesis window—leads to the loss of orthogonality in the transform. The question then arises as to whether perfect reconstruction can still be achieved. If proven possible, the resulting filter bank is known as *biorthogonal*, a term commonly used in the wavelet literature [87, 88, 89]. It would then be interesting to know what design constraints would be required to maintain biorthogonality. Towards this end, a time-domain analysis similar to the one performed by Malvar for the ELT is necessary.

4.2.1 Time Domain Analysis

Consider the infinite input vector $\tilde{\mathbf{x}}$, the infinite reconstructed vector $\tilde{\mathbf{y}}$ and the infinite transform vector $\tilde{\mathbf{X}}$. We have

$$\tilde{\mathbf{X}} = \tilde{\mathbf{P}}^T \tilde{\mathbf{x}} \quad (4.9)$$

and, if no processing is performed on the transform coefficients, we have

$$\tilde{\mathbf{y}} = \tilde{\mathbf{Q}}\tilde{\mathbf{X}} \quad (4.10)$$

where the infinite matrices $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$ are given by

$$\tilde{\mathbf{P}} \equiv \begin{pmatrix} \ddots & & & & 0 \\ & \mathbf{P}_0 & & & \\ & \mathbf{P}_1 & \mathbf{P}_0 & & \\ & \mathbf{P}_2 & \mathbf{P}_1 & \mathbf{P}_0 & \\ & \vdots & \vdots & \vdots & \\ & \mathbf{P}_{2K-1} & \mathbf{P}_{2K-2} & \mathbf{P}_{2K-3} & \\ & & \mathbf{P}_{2K-1} & \mathbf{P}_{2K-2} & \\ & & & \mathbf{P}_{2K-1} & \\ 0 & & & & \ddots \end{pmatrix}, \quad \tilde{\mathbf{Q}} \equiv \begin{pmatrix} \ddots & & & & 0 \\ & \mathbf{Q}_0 & & & \\ & \mathbf{Q}_1 & \mathbf{Q}_0 & & \\ & \mathbf{Q}_2 & \mathbf{Q}_1 & \mathbf{Q}_0 & \\ & \vdots & \vdots & \vdots & \\ & \mathbf{Q}_{2K-1} & \mathbf{Q}_{2K-2} & \mathbf{Q}_{2K-3} & \\ & & \mathbf{Q}_{2K-1} & \mathbf{Q}_{2K-2} & \\ & & & \mathbf{Q}_{2K-1} & \\ 0 & & & & \ddots \end{pmatrix}.$$

Combining equations (4.9) and (4.10) we get

$$\tilde{\mathbf{y}} = \tilde{\mathbf{Q}}\tilde{\mathbf{P}}^T \tilde{\mathbf{x}}$$

where perfect reconstruction requires $\tilde{\mathbf{y}} = \tilde{\mathbf{x}}$. This is achieved if and only if

$$\tilde{\mathbf{Q}}\tilde{\mathbf{P}}^T = \mathbf{I} \quad \iff \quad \sum_{m=\max(0,-l)}^{\min(2K-1,2K-1-l)} \mathbf{Q}_m \mathbf{P}_{m+l}^T = \delta(l)\mathbf{I}, \quad \text{for } l = -(2K-1), \dots, 2K-1. \quad (4.11)$$

This is the *biorthonormal condition* [58].

To apply condition (4.11) to the context of selection of the analysis and synthesis windows requires the introduction of more notation. The following serves mostly to isolate the window functions $h[n]$ and $f[n]$ from the transform matrices. First, we define the square analysis-window matrix \mathbf{H} of order $2KM$ by

$$\mathbf{H} \equiv \text{diag}\{h[0], h[1], \dots, h[2KM-1]\} \equiv \text{diag}\{\mathbf{H}_0, \mathbf{H}_1, \dots, \mathbf{H}_{2K-1}\}$$

where \mathbf{H}_l is the l th diagonal square block of order M as in

$$\mathbf{H}_l \equiv \text{diag}\{h[lM], h[lM+1], \dots, h[lM+M-1]\}.$$

The synthesis-window matrix \mathbf{F} can also be defined in a similar manner.

$$\mathbf{F} \equiv \text{diag}\{f[0], f[1], \dots, f[2KM - 1]\} \equiv \text{diag}\{\mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_{2K-1}\}$$

where \mathbf{F}_l is the l th diagonal square block of order M as in

$$\mathbf{F}_l \equiv \text{diag}\{f[lM], f[lM + 1], \dots, f[lM + M - 1]\}.$$

Second, we define the $2KM$ by M modulation matrix Φ which has element $[\Phi]_{nk}$ given by

$$[\Phi]_{nk} \equiv \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right].$$

Again, we subdivide Φ into square blocks Φ_l of order M , such that

$$\Phi^T \equiv [\Phi_0^T \Phi_1^T \dots \Phi_{2K-1}^T].$$

It is then easy to verify that the transform matrices as defined in (4.4) and (4.5) can be written as

$$\mathbf{P} = \mathbf{H}\Phi \quad \text{and} \quad \mathbf{P}_l = \mathbf{H}_l\Phi_l \quad \text{for } l = 0, 1, \dots, 2K - 1 \quad (4.12)$$

$$\mathbf{Q} = \mathbf{F}\Phi \quad \text{and} \quad \mathbf{Q}_l = \mathbf{F}_l\Phi_l \quad \text{for } l = 0, 1, \dots, 2K - 1. \quad (4.13)$$

We can also verify, after some algebraic manipulation (refer to [86] for details), that concatenating the forward and inverse operations of the modulation matrices will lead to the following aliasing.

$$\begin{cases} \Phi_i \Phi_{i+2s}^T & = (-1)^s [\mathbf{I} + (-1)^i \mathbf{J}] \\ \Phi_i \Phi_{i+2s+1}^T & = 0 \end{cases} \quad \text{for } \begin{cases} i & = 0, 1, \dots, 2K - 1 \\ s & = 0, 1, \dots, 2K - 1 - 2i \end{cases} \quad (4.14)$$

where \mathbf{J} is the counter-identity matrix, given by (4.3).

We now attempt to simplify the biorthonormality condition in (4.11) to form a condition on the analysis and synthesis windows by substituting (4.12) and (4.13) into (4.11) and using (4.14):

$$\begin{aligned} \sum_{m=\max(0,-l)}^{\min(2K-1,2K-1-l)} \mathbf{Q}_m \mathbf{P}_{m+l}^T &= \sum_{m=\max(0,-l)}^{\min(2K-1,2K-1-l)} \mathbf{F}_m \Phi_m \Phi_{m+l}^T \mathbf{H}_{m+l}^T && \text{[set } l = 2s\text{]} \\ &= \begin{cases} \sum_{m=\max(0,-2s)}^{\min(2K-1,2K-1-2s)} (-1)^s \mathbf{F}_m [\mathbf{I} + (-1)^m \mathbf{J}] \mathbf{H}_{m+2s}^T & \text{for integer } s \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Using this, (4.11) can therefore be simplified into the following two equations:

$$\sum_{m=\max(0,-2s)}^{\min(2K-1,2K-1-2s)} \mathbf{F}_m \mathbf{H}_{m+2s} = \delta(s) \mathbf{I} \quad \text{for } s = -(K-1), \dots, K-1 \quad (4.15)$$

$$\sum_{m=\max(0,-2s)}^{\min(2K-1,2K-1-2s)} (-1)^m \mathbf{F}_m \mathbf{J} \mathbf{H}_{m+2s} = 0 \quad \text{for } s = -(K-1), \dots, K-1. \quad (4.16)$$

These are the preliminary constraints on the analysis and synthesis windows in matrix form. In the next section we will simplify these constraints.

4.2.2 Simplification of the Constraints

The matrix equations in (4.15) and (4.16), while compact, offer very little insight. In this particular case, it is better to reduce the equations to scalar form. Towards that end, we first reduce the number of matrix equations. Observe that each equation in (4.15) and (4.16) can be reduced to two groups of matrix equations according to the sign of s . Working with (4.15) first, so we have

$$\sum_{m=0}^{2K-1-2s} \mathbf{F}_m \mathbf{H}_{m+2s} = \delta(s) \mathbf{I} \quad \text{for } s = 0, \dots, K-1 \quad (4.17)$$

$$\sum_{m=-2s}^{2K-1} \mathbf{F}_m \mathbf{H}_{m+2s} = 0 \quad \text{for } s = -(K-1), \dots, -1. \quad (4.18)$$

Given that the analysis window $h[n]$ and synthesis window $f[n]$ are both selected to be symmetric, the window matrices, \mathbf{H} and \mathbf{F} , have the following property:

$$h[n] = h[2KM - 1 - n] \quad \iff \quad \mathbf{HJ} = \mathbf{JH} \quad \iff \quad \mathbf{JH}_l = \mathbf{H}_{2K-l-1} \mathbf{J}, \quad (4.19)$$

$$f[n] = f[2KM - 1 - n] \quad \iff \quad \mathbf{FJ} = \mathbf{JF} \quad \iff \quad \mathbf{JF}_l = \mathbf{F}_{2K-l-1} \mathbf{J}. \quad (4.20)$$

Using these properties, the matrix equations in (4.18) can be shown to be equivalent to the ones in (4.17).

$$\begin{aligned}
& \sum_{m=-2s}^{2K-1} \mathbf{F}_m \mathbf{H}_{m+2s} = 0 && \text{for } s = -(K-1), \dots, -1 \\
\iff & \sum_{m=-2s}^{2K-1} \mathbf{F}_m \mathbf{J} \mathbf{J} \mathbf{H}_{m+2s} = 0 && \text{[note that } \mathbf{J} \mathbf{J} = \mathbf{I}] \\
\iff & \sum_{m=-2s}^{2K-1} \mathbf{J} \mathbf{F}_{2K-1-m} \mathbf{H}_{2K-1-m-2s} \mathbf{J} = 0 && \text{[using (4.19) \& (4.20)]} \\
\iff & \sum_{m=-2s}^{2K-1} \mathbf{F}_{2K-1-m} \mathbf{H}_{2K-1-m-2s} = 0
\end{aligned}$$

Set $m' = 2K - 1 - m$ (note change in summation limits)

$$\iff \sum_{m'=0}^{2K-1+2s} \mathbf{F}_{m'} \mathbf{H}_{m'-2s} = 0$$

Set $s' = -s$

$$\iff \sum_{m'=0}^{2K-1-2s'} \mathbf{F}_{m'} \mathbf{H}_{m'+2s'} = 0 \quad \text{for } s' = 1, \dots, K-1.$$

We can see that this reduces to the same condition as in (4.17). A similar procedure can be applied to (4.16). The end result is the following two conditions.

$$\sum_{m=0}^{2K-1-2s} \mathbf{F}_m \mathbf{H}_{m+2s} = \delta(s) \mathbf{I} \quad \text{for } s = 0, \dots, K-1 \quad (4.21)$$

$$\sum_{m=0}^{2K-1-2s} (-1)^m \mathbf{F}_m \mathbf{J} \mathbf{H}_{m+2s} = 0 \quad \text{for } s = 0, \dots, K-1 \quad (4.22)$$

Therefore, the design constraints on the analysis and synthesis windows are reduced to the $2K$ matrix equations in (4.21) and (4.22). Each of these matrix equations corresponds to M scalar equations as in the following:

Design Constraints on the Analysis and Synthesis Windows

$$\sum_{m=0}^{2K-1-2s} f[mM+n]h[(m+2s)M+n] = \delta(s) \quad (4.23)$$

$$\sum_{m=0}^{2K-1-2s} (-1)^m f[mM+n]h[(m+2s)M+(M-n-1)] = 0 \quad (4.24)$$

for $s = 0, \dots, K-1$ and $n = 0, \dots, M-1$.

Advantages from Incorporation of Biorthogonality

To many engineers, the *raison d'être* of any theoretical development is that it eventually leads to practical improvement. For this reason, we explore in this chapter the advantages of incorporating biorthogonality into lapped transforms and investigate how the new filter bank can be used to improve audio coder performance. Relaxation of the orthogonality requirement in the transform should lead to more degrees of design freedom. This additional freedom is important because it directly affects the ability of the designer to create filter banks with desired properties. Therefore, an important question is to quantify the increase in degrees of freedom obtained by allowing biorthogonality in the filter bank. In the following, we address this question for three special cases ($K = 1, 2, 3$). A general result is then shown and proven. The chapter concludes with a suggestion for how this increased flexibility can be valuable in audio compression.

5.1 Increase in Degrees of Freedom: Special Cases

In the previous chapter, we arrived at the equations in (4.23) and (4.24), conditions required to achieve perfect reconstruction in the biorthogonal filter bank. Note that there is a total of $2KM$ design variables in the construction of the symmetric windows of the filter bank. If there are, as the number of equations indicate, $2KM$ conditions, then we would have no design freedom. Fortunately, unlike the ELT constraints in which the $KM/2$ equations are all independent, there is a certain amount of dependency among the $2KM$ equations for the biorthogonal case. In the following we will endeavor to remove the dependencies in order to determine exactly how many degrees of freedom we have gained from incorporating biorthogonality.

5.1.1 Special Case I: $K = 1$

When the overlapping factor is one, only adjacent transform frames overlap. In lemmas 5.1 and 5.2, we reduce the equations that are associated with $s = 0$. These lemmas will be useful in subsequent

special cases. The theorem that follows will establish the reduced set of equations for the $K = 1$ case. The resulting number of degrees of freedom for this case is given in the corollary.

Bear in mind throughout the mathematical development in this chapter that the analysis window $h[n]$ and the synthesis window $f[n]$ are both symmetric as indicated in (4.7). We will also assume that all the relevant coefficients are non-zero:

$$f[n] \neq 0, \quad h[n] \neq 0, \quad \text{for } n = 0, \dots, 2KM - 1.$$

LEMMA 5.1 (Reduction of first set of equations for $s = 0$)

In the following set of M equations,

$$\sum_{m=0}^{2K-1} f[mM + n]h[mM + n] = 1 \quad \text{for } n = 0, \dots, M - 1,$$

the equations for $n = n_0$ are equivalent to the equations for $n = M - 1 - n_0$.

Proof.

$$\begin{aligned} & \sum_{m=0}^{2K-1} f[mM + n_0]h[mM + n_0] = 1 \\ \Leftrightarrow & \sum_{m=0}^{2K-1} f[2KM - mM - n_0 - 1]h[2KM - mM - n_0 - 1] = 1 \\ \Leftrightarrow & \sum_{m=0}^{2K-1} f[(2K - 1 - m)M + (M - n_0 - 1)]h[(2K - 1 - m)M + (M - n_0 - 1)] = 1 \end{aligned}$$

set $m' = 2K - 1 - m$ (no change in summation limits)

$$\Leftrightarrow \sum_{m=0}^{2K-1} f[m'M + (M - n_0 - 1)]h[m'M + (M - n_0 - 1)] = 1$$

□

LEMMA 5.2 (Reduction of 2nd set of equations for $s = 0$)

In the following set of M equations,

$$\sum_{m=0}^{2K-1} (-1)^m f[mM + n]h[(m + 1)M - n - 1] = 1 \quad \text{for } n = 0, \dots, M - 1,$$

the equations for $n = n_0$ are equivalent to the equations for $n = M - 1 - n_0$.

Proof.

$$\begin{aligned} & \sum_{m=0}^{2K-1} (-1)^m f[mM + n_0]h[(m+1)M - n_0 - 1] = 0 \\ \iff & \sum_{m=0}^{2K-1} (-1)^m f[2KM - mM - n_0 - 1]h[2KM - mM - M - n_0] = 0 \\ \iff & \sum_{m=0}^{2K-1} (-1)^m f[(2K-1-m)M + (M - n_0 - 1)]h[(2K-1-m)M - n_0] = 0 \end{aligned}$$

set $m' = 2K - 1 - m$ (no change in summation limits)

$$\iff \sum_{m'=0}^{2K-1} (-1)^{m'} f[m'M + (M - n_0 - 1)]h[(m'+1)M - (M - n_0 - 1) - 1] = 0$$

□

THEOREM 5.3 (Design constraints in $K = 1$ case)

The constraints on analysis window $h[n]$ and synthesis window $f[n]$, required for perfect reconstruction in the biorthogonal cosine-modulated filter bank, reduce, in the $K = 1$ case, to the following set of equations:

$$f[n]h[n] + f[n+M]h[n+M] = 1 \quad (5.1)$$

$$f[n]h[n+M] - f[n+M]h[n] = 0 \quad (5.2)$$

where $n = 0, \dots, \frac{M}{2} - 1$.

Proof. When $K = 1$, we only have one set of equations corresponding to $s = 0$, therefore, by simple substitution, we get the following:

$$M \text{ equations in (4.23)} \xrightarrow{K=1, s=0} \text{Equations in (5.1) for } n = 0, \dots, M-1$$

$$M \text{ equations in (4.24)} \xrightarrow{K=1, s=0} \text{Equations in (5.2) for } n = 0, \dots, M-1.$$

From lemmas 5.1 and 5.2, we know that, in both cases, $M/2$ of the equations are equivalent to the other $M/2$. Therefore the range of n for the above equations can be reduced from $[0, M)$ to $[0, \frac{M}{2})$.

This gives the desired result. \square

COROLLARY 5.4 (Degrees of freedom in $K = 1$ case)

For the $K = 1$ case, there are at least M degrees of freedom in choosing the combined total of $2M$ coefficients in the analysis and synthesis windows in the biorthogonal cosine-modulated filter bank.

Proof. This follows directly from theorem 5.3. There are $\frac{M}{2} + \frac{M}{2} = M$ design constraints on the $2M$ design variables. This leaves M degrees of freedom. \square

An equivalent result has been observed in the context of time-domain aliasing cancellation [90].

5.1.2 Special Case II: $K = 2$

This case is slightly more complicated than the previous one, but similar arguments can be made to reduce the constraints. First, it is necessary to introduce lemma 5.5 in which the number of design constraints which are originally associated with the ELT is reduced. Lemma 5.5 will also be useful in latter cases.

LEMMA 5.5 (Reduction of the ELT design constraints)

In the following set of equations, for all $K, M,$

$$\sum_{m=0}^{2K-1-2s} h[mM+n]h[(m+2s)M+n] = \delta(s) \quad \text{where} \quad \begin{cases} n = 0, \dots, M-1, \\ s = 0, \dots, K-1, \end{cases}$$

the equations for $n = n_0$ are equivalent to the equations for $n = M - 1 - n_0$.

Proof.

$$\begin{aligned} & \sum_{m=0}^{2K-1-2s} h[mM+n_0]h[(m+2s)M+n_0] = \delta(s) \\ \Leftrightarrow & \sum_{m=0}^{2K-1-2s} h[2KM - mM - n_0 - 1]h[2KM - (m+2s)M - n_0 - 1] = \delta(s) \\ \Leftrightarrow & \sum_{m=0}^{2K-1-2s} h[(2K-1-m)M + (M-n_0-1)] \\ & \cdot h[(2K-1-m-2s)M + (M-n_0-1)] = \delta(s) \end{aligned}$$

Set $m' = 2K - 1 - 2s - m$ (no change in summation limits)

$$\iff \sum_{m'=0}^{2K-1-2s} h[m'M + (M - n_0 - 1)]h[(m' + 2s)M + (M - n_0 - 1)] = \delta(s)$$

□

LEMMA 5.6 (Reduction of set of equations for $s = K - 1$)

Given the set of M equations,

$$\sum_{m=0}^1 (-1)^m f[mM + n]h[(m + 2K - 2)M + (M - n - 1)] = 0 \quad \text{for } n = 0, \dots, M - 1, \quad (5.3)$$

then the set of M equations,

$$\sum_{m=0}^1 f[mM + n]h[(m + 2K - 2)M + n] = 0 \quad \text{for } n = 0, \dots, M - 1, \quad (5.4)$$

is equivalent to the following set of $M/2$ equations,

$$\sum_{m=0}^1 h[mM + n]h[(m + 2K - 2)M + n] = 0 \quad \text{for } n = 0, \dots, \frac{M}{2} - 1. \quad (5.5)$$

Proof. Note first that (5.3) can be written as follows:

$$\begin{aligned} & f[n]h[(2K - 1)M - n - 1] - f[M + n]h[2KM - n - 1] = 0 \\ \iff & f[n]h[(2K - 1)M - n - 1] = f[M + n]h[2KM - n - 1] \end{aligned} \quad (5.6)$$

We can then show that equation (5.4)

$$\begin{aligned} \iff & f[n]h[(2K - 2)M + n] \\ & \quad + f[M + n]h[(2K - 1)M + n] = 0 \\ \iff & f[n]h[(2K - 1)M - n - 1]h[(2K - 2)M + n] \\ & \quad + f[M + n]h[(2K - 1)M - n - 1]h[(2K - 1)M + n] = 0 \\ \iff & f[M + n]h[2KM - n - 1]h[(2K - 2)M + n] \\ & \quad + f[M + n]h[(2K - 1)M - n - 1]h[(2K - 1)M + n] = 0 \quad \text{[by (5.6)]} \\ \iff & h[2KM - n - 1]h[(2K - 2)M + n] \end{aligned}$$

$$\begin{aligned}
& +h[(2K-1)M-n-1]h[(2K-1)M+n] = 0 \\
\iff & h[n]h[(2K-2)M+n] + h[n+M]h[(2K-1)M+n] = 0 \quad [\text{by (4.7)}] \\
\iff & \hspace{10em} (5.5) \hspace{10em} \text{for } n = 0, \dots, M-1 \quad (5.7)
\end{aligned}$$

The set of equations in (5.7) are for $n \in [0, M)$ and therefore are not completely equivalent to those in (5.5). However, since (5.7) is the exact same set of equations as the ELT design constraints, invoking lemma 5.5 reduces the range of n to $[0, \frac{M}{2})$ as desired. \square

THEOREM 5.7 (Design constraints in $K = 2$ case)

The constraints on analysis window $h[n]$ and synthesis window $f[n]$, required for perfect reconstruction in the biorthogonal cosine-modulated filter bank, reduces, in the $K = 2$ case, to the following set of equations.

$$f[n]h[n] + f[M+n]h[M+n] \tag{5.8}$$

$$+ f[2M+n]h[2M+n] + f[3M+n]h[3M+n] = 1$$

$$f[n]h[3M+n] - f[M+n]h[2M+n] \tag{5.9}$$

$$+ f[2M+n]h[M+n] - f[3M+n]h[n] = 0$$

$$f[n]h[M+n] - f[M+n]h[n] = 0 \tag{5.10}$$

$$h[n]h[2M+n] + h[M+n]h[3M+n] = 0 \tag{5.11}$$

for $n = 0, \dots, \frac{M}{2} - 1$ in (5.8), (5.9) and (5.11), and $n = 0, \dots, M-1$ in (5.10).

Proof. The equations associated with $s = 0$ can be reduced for the $K = 2$ case in the same way as the $K = 1$ case using lemmas 5.1 and 5.2 (see proof of theorem 5.3) as follows:

$$M \text{ equations in (4.23)} \xrightarrow{K=2, s=0} M/2 \text{ equations in (5.8),}$$

$$M \text{ equations in (4.24)} \xrightarrow{K=2, s=0} M/2 \text{ equations in (5.9).}$$

The equations associated with $s = K - 1 = 1$ can be reduced with the result from lemma 5.6 as follows:

$$\left\{ \begin{array}{l} M \text{ equations in (4.23)} \\ M \text{ equations in (4.24)} \end{array} \right. \xrightarrow{K=2, s=1} \left\{ \begin{array}{l} M \text{ equations in (5.10)} \\ M/2 \text{ equations in (5.11).} \end{array} \right.$$

\square

COROLLARY 5.8 (Degrees of freedom in $K = 2$ case)

For the $K = 2$ case, there are at least $3M/2$ degrees of freedom in choosing the combined total of $4M$ coefficients in the analysis and synthesis windows in the biorthogonal cosine-modulated filter bank.

Proof. This follows from theorem 5.7. We can count a total of $5M/2$ design constraints for the $4M$ design variables. That leaves $3M/2$ degrees of freedom. \square

On comparison with corollary 5.4, the latest result comes as somewhat of a surprise. The doubling of degrees of freedom for the $K = 1$ case is fairly intuitive and it suggests similar results for higher K . Instead, at $K = 2$, we already see that the increase in degrees of freedom has fallen short of expectations. It is therefore instructive to study at least one more special case of K .

5.1.3 Special Case III: $K = 3$

As in the previous two cases, we begin with lemma 5.9 in which we reduce the number of equations associated with $s = K - 2$.

LEMMA 5.9 (Reduction of set of equations for $s = K - 2$)

Given the following sets of equations,

$$\sum_{m=0}^1 f[mM + n]h[(m + 2K - 2)M + n] = 0 \quad (5.12)$$

$$\sum_{m=0}^1 (-1)^m f[mM + n]h[(m + 2K - 2)M + (M - n - 1)] = 0 \quad (5.13)$$

$$\sum_{m=0}^3 (-1)^m f[mM + n]h[(m + 2K - 4)M + (M - n - 1)] = 0 \quad (5.14)$$

all for $n = 0, \dots, M - 1$, then the set of equations,

$$\sum_{m=0}^3 f[mM + n]h[(m + 2K - 4)M + n] = 0 \quad \text{for } n = 0, \dots, M - 1, \quad (5.15)$$

is equivalent to

$$\sum_{m=0}^3 h[mM + n]h[(m + 2K - 4)M + n] = 0 \quad \text{for } n = 0, \dots, \frac{M}{2} - 1. \quad (5.16)$$

Proof. First we rewrite the equations in (5.12), (5.13) and (5.14), using the symmetric property in (4.7),

$$(5.12) \quad \iff f[n]h[n+M] = f[n+M]h[n] \quad (5.17)$$

$$(5.13) \quad \iff f[n]h[(2K-2)M+n] = -f[n+M]h[(2K-1)M+n] \quad (5.18)$$

$$(5.14) \quad \iff \begin{aligned} & f[n+2M]h[n+M] - f[n+3M]h[n] \\ & = f[n+M]h[n+2M] - f[n]h[n+3M] \end{aligned} \quad (5.19)$$

all for $n = 0, \dots, M-1$. From lemma 5.6 we also know that, given (5.13), (5.12) is equivalent to the following:

$$\begin{aligned} & \sum_{m=0}^1 h[mM+n]h[(m+2K-2)M+n] = 0 \\ \iff & -h[n]h[(2K-2)M+n] = h[n+M]h[(2K-1)M+n]. \end{aligned} \quad (5.20)$$

Based on the above, we can then show that the equations from (5.16)

$$\begin{aligned} \iff & f[n]h[(2K-4)M+n] + f[M+n]h[(2K-3)M+n] \\ & + f[2M+n]h[(2K-2)M+n] \\ & + f[3M+n]h[(2K-1)M+n] = 0 \\ \iff & h[M+n]f[n]h[(2K-4)M+n] \\ & + h[M+n]f[M+n]h[(2K-3)M+n] \\ & + h[M+n]f[2M+n]h[(2K-2)M+n] \\ & + h[M+n]f[3M+n]h[(2K-1)M+n] = 0 \\ \iff & f[M+n]h[n]h[(2K-4)M+n] \\ & + f[M+n]h[M+n]h[(2K-3)M+n] \\ & + h[M+n]f[2M+n]h[(2K-2)M+n] \\ & - f[3M+n]h[n]h[(2K-2)M+n] = 0 \quad [\text{using (5.17) \& (5.20)}] \\ \iff & f[M+n]h[n]h[(2K-4)M+n] \\ & + f[M+n]h[M+n]h[(2K-3)M+n] \\ & + f[M+n]h[2M+n]h[(2K-2)M+n] \\ & - f[n]h[3M+n]h[(2K-2)M+n] = 0 \quad [\text{using (5.19)}] \\ \iff & f[M+n]h[n]h[(2K-4)M+n] \end{aligned}$$

$$\begin{aligned}
& +f[M+n]h[M+n]h[(2K-3)M+n] \\
& +f[M+n]h[2M+n]h[(2K-2)M+n] \\
& \quad +f[M+n]h[3M+n]h[(2K-1)M+n] = 0 \quad [\text{using (5.18)}] \\
\iff & h[n]h[(2K-4)M+n] + h[M+n]h[(2K-3)M+n] \\
& \quad +h[2M+n]h[(2K-2)M+n] \\
& \quad \quad +h[3M+n]h[(2K-1)M+n] = 0 \\
\iff & \quad \quad \quad (5.16) \quad \quad \quad \text{for } n = 0, \dots, M-1 \quad (5.21)
\end{aligned}$$

Invoking lemma 5.5 reduces the range of n for (5.21) from $[0, M)$ to $[0, \frac{M}{2})$ to match (5.16). \square

THEOREM 5.10 (Design constraints in $K = 3$ case)

The constraints on analysis window $h[n]$ and synthesis window $f[n]$, required for perfect reconstruction in the biorthogonal cosine-modulated filter bank, reduce, in the $K = 2$ case, to the following set of equations:

$$\sum_{m=0}^5 f[mM+n]h[mM+n] = 1 \quad (5.22)$$

$$\sum_{m=0}^5 f[mM+n]h[(m+1)M-n-1] = 1 \quad (5.23)$$

$$f[n]h[3M-n-1] + f[M+n]h[4M-n-1] \quad (5.24)$$

$$+f[2M+n]h[5M-n-1] + f[3M+n]h[6M-n-1] = 1$$

$$h[n]h[2M+n] + h[M+n]h[3M+n] \quad (5.25)$$

$$+h[2M+n]h[4M+n] + h[3M+n]h[5M+n] = 0$$

$$f[n]h[M+n] - f[M+n]h[n] = 0 \quad (5.26)$$

$$h[n]h[4M+n] + h[M+n]h[5M+n] = 0 \quad (5.27)$$

for $n = 0, \dots, \frac{M}{2} - 1$ in (5.22), (5.23), (5.25), and (5.27), and $n = 0, \dots, M-1$ in (5.24) and (5.26).

Proof. The equations associated with $s = 0$ can be reduced for the $K = 3$ case in the same way as the $K = 1$ and $K = 2$ cases using lemmas 5.1 and 5.2 (see proof of theorem 5.3):

$$M \text{ equations in (4.23)} \xrightarrow{K=3, s=0} M/2 \text{ equations in (5.22)}$$

$$M \text{ equations in (4.24)} \xrightarrow{K=3, s=0} M/2 \text{ equations in (5.23).}$$

The equations associated with $s = K - 1 = 2$ can be reduced with the result from lemma 5.6 (see proof of theorem 5.7):

$$\left\{ \begin{array}{l} M \text{ equations in (4.23)} \\ M \text{ equations in (4.24)} \end{array} \right. \xrightarrow{K=3, s=2} \left\{ \begin{array}{l} M \text{ equations in (5.26)} \\ M/2 \text{ equations in (5.27)}. \end{array} \right.$$

The equations associated with $s = K - 2 = 1$ can be reduced with the result from lemma 5.9. Taking the equations associated with $s = 2$ as given,

$$\left\{ \begin{array}{l} M \text{ equations in (4.23)} \\ M \text{ equations in (4.24)} \end{array} \right. \xrightarrow{K=3, s=1} \left\{ \begin{array}{l} M \text{ equations in (5.24)} \\ M/2 \text{ equations in (5.25)}. \end{array} \right.$$

□

COROLLARY 5.11 (Degrees of Freedom in $K = 3$ case)

For the $K = 3$ case, there are at least $2M$ degrees of freedom in choosing the combined total of $6M$ coefficients in the analysis and synthesis windows in the biorthogonal cosine-modulated filter bank.

Proof. This follows from theorem 5.10. There is a total of $4M$ design constraints on $6M$ design variables leaving $2M$ degrees of freedom. □

5.2 Increase in Degrees of Freedom: General Case

5.2.1 Emerging Pattern from Special Cases

The results from the special cases in the previous section are summarized in table 5.1. At this point, a pattern emerges. Under lemmas 5.1 and 5.2, the constraints associated with $s = 0$ are reduced from $2M$ to M equations. For the constraints associated with $s > 0$, however, the reduction is from $2M$ to $3M/2$ equations. We can also see that lemma 5.6 will be applicable to the constraints associated with $s = K - 1$ for all K , and the same can be said of lemma 5.9 for $s = K - 2$.

It is therefore not hard to extrapolate to the case of $K = 4$. The constraints associated with $s = 0, 2, 3$ will be reduced using the lemmas 5.1, 5.2, 5.6 and 5.9 as before. We can further conjecture the existence of another lemma (for the case $s = K - 3$) which will reduce the $2M$ equations associated with $s = 1$ to $M/2$ equations. This will give a total of $11M/2$ constraints for $8M$ design variables, leaving $5M/2$ degrees of freedom. Extending this conjecture to all K , we can expect $(3K - 1)M/2$ constraints for $2KM$ variables, leaving $(K + 1)M/2$ degrees of freedom. If

Table 5.1: Reduction of constraints for different values of K and s , with associated lemmas.

	$K = 1$	$K = 2$	$K = 3$
$s = 0$	$2M \rightarrow M$ equations lemmas 5.1 & 5.2	$2M \rightarrow M$ equations lemmas 5.1 & 5.2	$2M \rightarrow M$ equations lemmas 5.1 & 5.2
$s = 1$	N/A	$2M \rightarrow 3M/2$ equations lemma 5.6 for $s = K - 1$	$2M \rightarrow 3M/2$ equations lemma 5.9 for $s = K - 2$
$s = 2$		N/A	$2M \rightarrow 3M/2$ equations lemma 5.6 for $s = K - 1$
No. of Variables	$2M$	$4M$	$6M$
No. of Equations	M	$5M/2$	$4M$
Degrees of Freedom	M	$3M/2$	$2M$

proven true, this represents an increase of $M/2$ degrees of freedom over the orthogonal case (ELT).

Table 5.2: Conjecture for the special case of $K = 4$ and the general case for all K .

	$K = 4$	all K
No. of Variables	$8M$	$2KM$
No. of Equations	$11M/2$	$(3K - 1)M/2$
Degrees of Freedom	$5M/2$	$(K + 1)M/2$

The form of the reduced design constraints can also be extrapolated from the special cases. In general, the design constraints fall into three groups.

- *Group (1)*

$$\sum_{m=0}^{2K-1-2s} h[mM + n]h[(m + 2s)M + n] = 0$$

where $s = 1, \dots, K - 1$ and $n = 0, \dots, \frac{M}{2} - 1$. Note that these are the exact same constraints as in the orthogonal case, except for the equations for $s = 0$ which are removed. The removal of those $M/2$ equations accounts for the increase of $M/2$ degrees of freedom.

- *Group (2)*

$$\sum_{m=0}^{2K-1-2s} (-1)^m f[mM + n]h[(m + 2s)M - n - 1] = 0$$

where $s = 1, \dots, K - 1$ and $n = 0, \dots, M - 1$. The bilinear equations here will become linear

if one of the two windows is known. This is very important to the design process. Note again that the $s = 0$ case is not present.

- Group (3)

$$\sum_{m=0}^{2K-1} f[mM+n]h[mM+n] = 1$$

$$\sum_{m=0}^{2K-1} (-1)^m f[mM+n]h[(m+1)M-n-1] = 1$$

where $n = 0, \dots, \frac{M}{2} - 1$. This is the $s = 0$ case that is missing from equation group (2).

Notice that there are $(K-1)M$ equations in group (2) and M equations in group (3). Assuming that the conjecture is correct and the analysis window $h[n]$ is accurately designed, the equations in groups (2) and (3) combine to become KM linear constraints on the KM variables of the synthesis window $f[n]$. In other words, given $h[n]$, $f[n]$ will be uniquely determined by these KM equations.

The above conjecture was empirically tested for cases up to $K = 4$ by using randomly selected analysis windows that conform to the constraints in group (1). The empirical tests show that perfect reconstruction is always achieved. Furthermore, the synthesis window is always uniquely determined, which is a good indication that the equations in groups (2) and (3) are independent. Since we also know that the constraints in group (1) are the same as the ELT constraints, they should also be independent. Therefore, we are fairly certain that the total of $(3K-1)M/2$ equations represent independent constraints.

5.2.2 Theorem & Proof for the General Case

In this section, we attempt to prove the conjecture detailed in the previous section. In theorem 5.12, we will also present and prove the closed form solution for $f[n]$, given $h[n]$ is correctly designed and known.

THEOREM 5.12 (Independent Design Constraints)

The constraints on analysis window $h[n]$ and synthesis window $f[n]$, required for perfect reconstruction in the biorthogonal cosine-modulated filter bank reduces to the following set of equations:

$$\sum_{m=0}^{2K-1-2s} h[mM+n]h[(m+2s)M+n] = 0 \quad (5.28)$$

for $s = 1, \dots, K - 1$ and $n = 0, \dots, \frac{M}{2} - 1$, and

$$f[n] = \frac{h[n]}{\sum_{m=0}^{2K-1} (h[mM + (n \bmod M)])^2} \quad (5.29)$$

for $n = 0, \dots, KM - 1$.

Proof. This proof is separated into three parts. First, we analyze an alternative formulation of the biorthogonal filter bank. This formulation is basically the ELT in conjunction with an amplitude-equalization step. It will be shown that the design constraints on the ELT analysis window is the same as (5.28) and the rest of the system, namely the equalization, is uniquely determined. In the second step, we will show that, under perfect reconstruction, the alternative formulation is equivalent to the biorthogonal filter bank we have studied originally. The third step will allow us to incorporate the equalization into the synthesis window giving (5.29). It is then easy to show by substitution that (5.29) will satisfy the KM linear equations in groups (2) and (3) in the previous section.

Step 1 Alternative Formulation (ELT with amplitude equalization)

Figure 5.1 is a simplified version of figure 4.1 which shows a system based on a lapped transform. We will be using this simplified diagrammatic approach for our discussion.

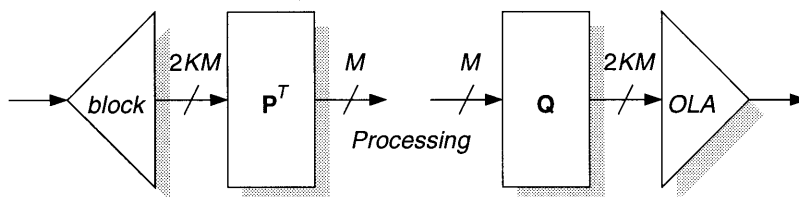


Figure 5.1: Simplified diagram of a system based on a lapped transform.

Now, consider the formulation in figure 5.2. This is basically the ELT with amplitude equalization—multiplication by $a[n]$ —after overlap-add is performed. Now, we would like to see what is necessary for this newly formulated system to satisfy perfect reconstruction. First we define $\tilde{\mathbf{B}}$ and $b[n]$ as follows

$$\tilde{\mathbf{B}} \equiv \text{diag}\{\dots, b[-1], b[0], b[1], \dots, b[n], \dots\} \quad \text{where} \quad b[n] = \frac{1}{a[n]}.$$

In the notation established in section 4.2.1, perfect reconstruction is achieved if and only if

$$\tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \tilde{\mathbf{B}} \quad (5.30)$$

where $\tilde{\mathbf{P}}$, to refresh your memory, is the infinite transform matrix.

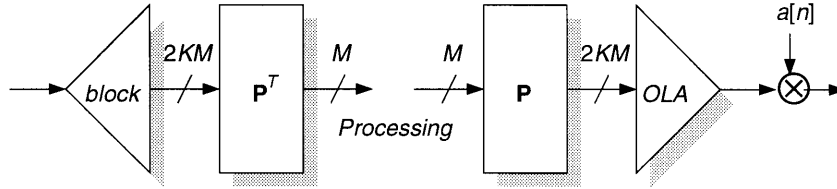


Figure 5.2: Alternative formulation of system (ELT with amplitude equalization).

As was evident from the time-domain analysis in section 4.1.2, $\tilde{\mathbf{P}}$ represents periodic application of the transform matrix \mathbf{P} . Under perfect reconstruction, $\tilde{\mathbf{B}}$ will also assume a periodic structure as in the following:

$$\tilde{\mathbf{B}} \equiv \text{diag}\{\dots, \mathbf{B}, \mathbf{B}, \mathbf{B}, \dots\}$$

where $\mathbf{B} \equiv \text{diag}\{b[0], b[1], \dots, b[M-1]\}$.

In other words, $b[n]$ is periodic with period M . The variation on the biorthonormal condition in (5.30) can then be rewritten as follows:

$$\sum_{m=0}^{2K-1-l} \mathbf{P}_m \mathbf{P}_{m+l}^T = \delta(l) \mathbf{B}, \quad \text{for } l = 0, \dots, 2K-1. \quad (5.31)$$

Again, the notation introduced in section 4.1.2 is used.

Using a similar technique as before, we simplify by separating the transform matrix into a window matrix and a modulation matrix:

$$\begin{aligned} \sum_{m=0}^{2K-1-l} \mathbf{P}_m \mathbf{P}_{m+l}^T &= \sum_{m=0}^{2K-1-l} \mathbf{H}_m \Phi_m \Phi_{m+l}^T \mathbf{H}_{m+l}^T && [\text{Set } l = 2s] \\ &= \begin{cases} \sum_{m=0}^{2K-1-2s} (-1)^s \mathbf{H}_m [\mathbf{I} + (-1)^m \mathbf{J}] \mathbf{H}_{m+2s}^T & \text{for integer } s \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Using this result, (5.31) can therefore be simplified into the following two equations:

$$\sum_{m=0}^{2K-1-2s} \mathbf{H}_m \mathbf{H}_{m+2s} = \delta(s) \mathbf{B} \quad \text{for } s = 0, \dots, K-1, \quad (5.32)$$

$$\sum_{m=0}^{2K-1-2s} (-1)^m \mathbf{H}_m \mathbf{J} \mathbf{H}_{m+2s} = 0 \quad \text{for } s = 0, \dots, K-1. \quad (5.33)$$

The left hand side of equation (5.33) is identically zero. Therefore, the only design constraints are given by (5.32), shown below in scalar form:

$$\sum_{m=0}^{2K-1-2s} h[mM+n]h[(m+2s)M+n] = \delta(s)b[n] \quad (5.34)$$

where $s = 0, \dots, K-1$ and $n = 0, \dots, M-1$. The M equations associated with $s = 0$ uniquely determine $b[n]$ but they do not constitute design constraints on $h[n]$. For $s = 1, \dots, K-1$, the number of equations can be reduced by half using lemma 5.5. We are therefore left with $(K-1)M/2$ constraints. Observe that the $(K-1)M/2$ constraints are the same as those in (5.28) and they lead to $(K+1)M/2$ degrees of freedom for designing $h[n]$.

Step 2 Equivalence between the original biorthogonal system and the alternative formulation

We return to the original formulation of the biorthogonal filter bank. Without loss of generality, we can rewrite the synthesis window $f[n]$ as follows:

$$f[n] = h[n]c[n] \quad \text{for } n = 0, \dots, M-1, \quad (5.35)$$

and in matrix notation, we can write

$$\mathbf{F} = \mathbf{H}\mathbf{C} \quad \text{and} \quad \mathbf{Q} = \mathbf{P}\mathbf{C}$$

where $\mathbf{C} \equiv \text{diag}\{c[0], c[1], \dots, c[2KM-1]\}$.

A diagrammatic representation is given in figure 5.3.

To show that the system as depicted in figure 5.3 is equivalent to that in figure 5.2, we have to show the two subsystems in figure 5.4 to be equivalent. That is to say, we require the use of post-overlap-add equalization $a[n]$ be equivalent to the use of pre-overlap-add equalization $c[n]$.

Given that $c[n]$ has $2K$ times more coefficients than $a[n]$, the two subsystems and therefore the two formulations will not be equivalent under most circumstances. As a matter of fact, it is

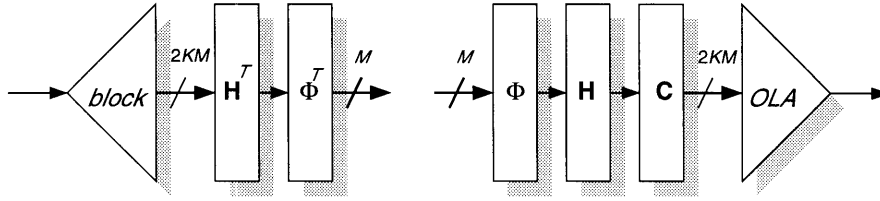
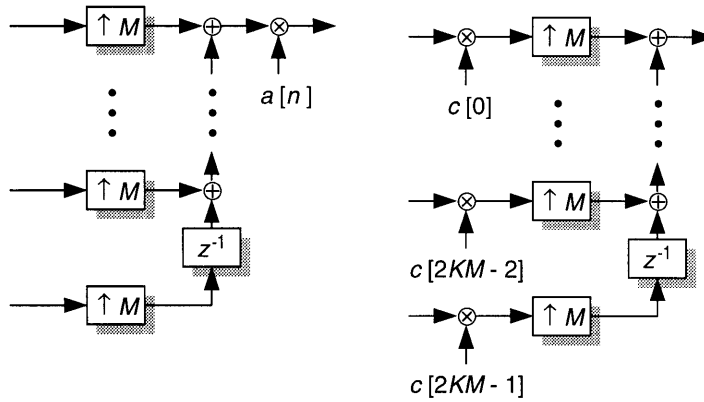


Figure 5.3: A different representation of the system in figure 5.1.



(a) From figure 5.2

(b) From figure 5.3

Figure 5.4: Subsystems from figures 5.2 and 5.3.

easy to verify that the two subsystems will be equivalent if and only if

$$c[n] = a[n \bmod M]. \tag{5.36}$$

In other words, $c[n]$ has to be periodic with period M . Fortunately, this is indeed the case when the biorthogonal filter bank satisfies perfect reconstruction. The design constraints in (4.23) and (4.24) in effect guarantee that $c[n]$ is periodic. Consider, for example, the case of $s = K - 1$. The constraint in (4.24) reduces to the following:

$$\begin{aligned} & f[n]h[n + M] - f[n + M]h[n] = 0 \\ \iff & c[n]h[n]h[n + M] - c[n + M]h[n]h[n + M] = 0 \\ \iff & c[n] = c[n + M]. \end{aligned} \tag{5.37}$$

Therefore, under perfect reconstruction, the original formulation of the biorthogonal filter bank is indeed equivalent to the alternative formulation studied in step 1.

Step 3 Formulation of synthesis window $f[n]$

In (5.34) of step (1), we have already shown that $b[n]$ is uniquely determined by $h[n]$ as follows:

$$b[n] = \sum_{m=0}^{2K-1} (h[n + mM])^2. \quad (5.38)$$

Combining this with (5.35) and (5.36) we get

$$\begin{aligned} f[n] &= c[n]h[n] = \frac{h[n]}{b[n \bmod M]} \\ &= \frac{h[n]}{\sum_{m=0}^{2K-1} (h[mM + (n \bmod M)])^2} \end{aligned} \quad (5.39)$$

for $n = 0, \dots, KM - 1$. It can be shown by substitution that these KM independent equations on $f[n]$ will satisfy all $2KM$ equations in (4.23) and (4.24). \square

COROLLARY 5.13 (Degrees of Freedom)

There are at least $(K+1)M/2$ degrees of freedom in choosing the combined total of $2KM$ coefficients in the analysis and synthesis windows in the biorthogonal cosine-modulated filter bank.

Proof. This follows from the previous proof. \square

5.2.3 Summary & Insight

Theorem 5.12 and the associated corollary 5.13 show that the previous conjecture is correct. For the design of a M -band biorthogonal cosine-modulated filter bank based on the ELT approach, we have $(K+1)M/2$ degrees of freedom. This represents an increase of $M/2$ degrees of freedom over the orthogonal case. In addition, the theorem also presents a closed-form solution for $f[n]$ given a correctly designed $h[n]$.

One interesting insight that can be gleaned from the proof to theorem 5.12 shows us where the increase of $M/2$ degrees of freedom comes from and also why it is limited to only $M/2$. It is possible to divide the perfect reconstruction requirement into two parts, alias cancellation and amplitude

equalization. By allowing biorthogonality while retaining the cosine-modulation structure of the lapped transform, we have only relaxed the amplitude-equalization part of the requirement; that is, the $M/2$ constraints associated with $s = 0$. The rest of the design constraints are still necessary to maintain alias cancellation. Therefore, the increase is limited to $M/2$ degrees of freedom.

5.3 Design of Biorthogonal Cosine-Modulated Filter Bank

The issue of filter bank design, while beyond the scope of this thesis, is nevertheless very important. In the biorthogonal cosine-modulated filter bank, only the analysis and the synthesis windows need to be designed. In this section, we outline two possible design approaches that will satisfy the constraints.

5.3.1 Design Method Based on Theorem

Theorem 5.12 points directly to an approach for designing the windows. We can first construct an analysis window $h[n]$ that satisfies the constraints in (5.28). Towards that end, algorithms for constrained optimization such as the augmented Lagrangian technique might be used [86]. Once $h[n]$ is known, $f[n]$ is uniquely determined by the equations in (5.29).

This design approach is exceptionally useful when $K = 1$. For the case of $K = 1$, there are no constraints associated with (5.28). We can therefore choose freely a desirable symmetric window $h[n]$, and then solve for $f[n]$ by using

$$f[n] = \frac{h[n]}{h^2[n] + h^2[n + M]}. \quad (5.40)$$

It should be noted that the symmetry of the system also allows us to design the synthesis window $f[n]$ first and then solve for the analysis window $h[n]$.

Having complete freedom in our choice of one of the windows can be very significant. In the orthogonal case, the design is subject to the $M/2$ constraints given in (4.8). This greatly restricts the ability of the designer to choose an appropriate window. Among the commonly-used windows, only the raised-sine window satisfies the constraints. It is given by

$$h[n] = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2M} \right]. \quad (5.41)$$

Further optimization is very difficult because the constraints are nonlinear. For the biorthogonal

filter bank, no similar condition exists for one of the two windows. It is possible, for example, to use the familiar Kaiser window in which the tradeoff between mainlobe width and sidelobe attenuation is controlled by a single parameter β . The Kaiser window is given by

$$h[n] = \frac{I_0[\beta(1 - [2(n - M/2)/M]^2)^{1/2}]}{I_0(\beta)} \quad (5.42)$$

where $I_0(\cdot)$ represents the zeroth order modified Bessel function of the first kind [91]. In figure 5.5, we have chosen Kaiser windows with different β as the analysis window. The pairs of analysis and synthesis windows shown are all valid and satisfy the perfect reconstruction requirement.

The drawback to this design method is that there is no direct control over the design of the second window, be it $h[n]$ or $f[n]$. Consider the design example in figure 5.5 again. As the sidelobe behavior in the frequency response of the analysis window improves with β , the corresponding synthesis window becomes increasingly misshapened and its frequency response deteriorates. However, as we shall see in section 5.4, this increased flexibility can still be of great use.

5.3.2 Joint-Analysis-and-Synthesis-Window Design

A potentially better method would be to design the analysis and synthesis windows jointly. For example, we can minimize the weighted stopband energies in a single error function:

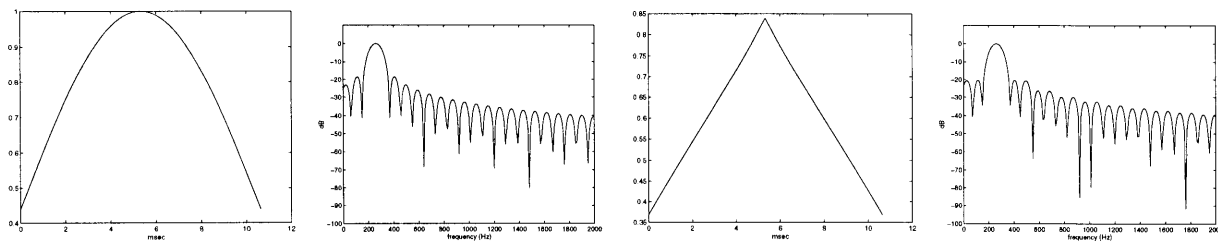
$$\epsilon = \frac{\alpha}{\pi} \int_{\omega_s}^{\pi} |H(e^{j\omega})|^2 d\omega + \frac{1 - \alpha}{\pi} \int_{\omega'_s}^{\pi} |F(e^{j\omega})|^2 d\omega \quad (5.43)$$

where ω_s and ω'_s are the stopband frequencies of the analysis filter and the synthesis filter respectively, and α is the weighting factor. This is of course subject to the design constraints either in the (4.23), (4.24) pair or in the (5.28), (5.29) pair. Again, constrained optimization methods such as the augmented Lagrangian technique can be used. Unfortunately, preliminary efforts have not led to a reasonable solution. Further study is required to make this design approach work.

5.4 Significance to Audio Compression

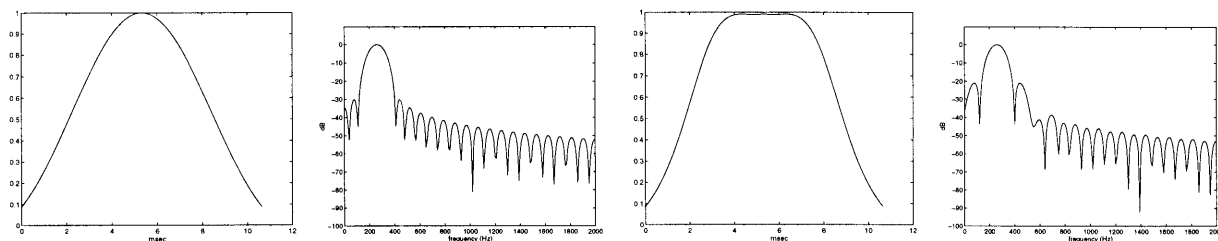
5.4.1 Filter-Bank Design Criteria in Audio Compression

To understand how the biorthogonal filter bank can be useful in audio compression schemes will require the re-examination of the design criteria detailed in section 2.2.1. The first two criteria,



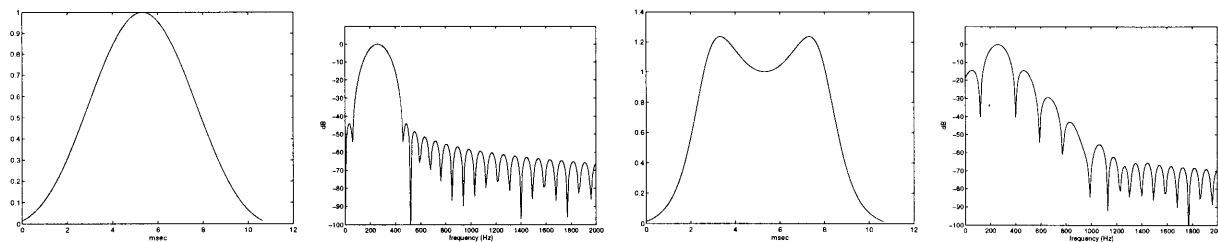
(a) Analysis Window (Kaiser $\beta = 2$)

(b) Corresponding Synthesis Window ($\beta = 2$)



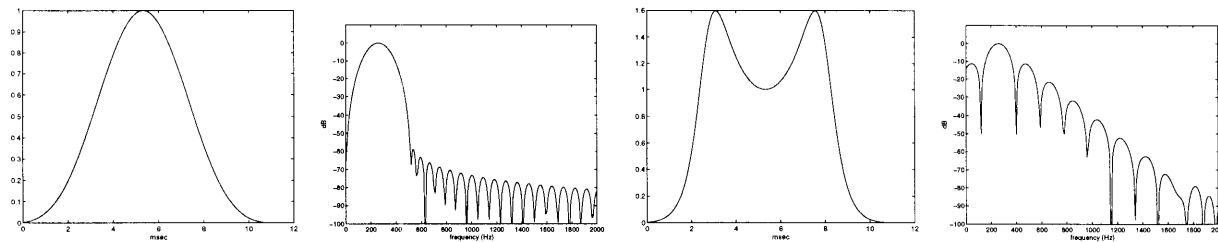
(c) Analysis Window (Kaiser $\beta = 4$)

(d) Corresponding Synthesis Window ($\beta = 4$)



(e) Analysis Window (Kaiser $\beta = 6$)

(f) Corresponding Synthesis Window ($\beta = 6$)



(g) Analysis Window (Kaiser $\beta = 8$)

(h) Corresponding Synthesis Window ($\beta = 8$)

Figure 5.5: Analysis and synthesis windows and their corresponding frequency responses for a special case ($K = 1$) of the biorthogonal cosine-modulated filter bank. 512-point (10.7 msec) Kaiser windows with $\beta = 2, 4, 6, 8$ are used as the analysis windows. Frequency responses are those of the windows modulated to 250 Hz.

perfect reconstruction and critical sampling, are automatically satisfied by the new formulation. The criterion of high frequency resolution, on the other hand, warrants further study. In the orthogonal case, the analysis and synthesis filters are identical. It is therefore not necessary to differentiate between the reasons for needing high frequency resolution in the two stages. In the biorthogonal case, however, it is useful to do so.

On the analysis side, as mentioned before, high frequency resolution is required for transform coding gain and perceptual modeling. On the synthesis side, however, what is more important is for the synthesis filters or basis functions to have minimum leakage into other subbands, or, in other words, to have good sidelobe behavior. Recall that in a typical audio coder, a masking curve is derived from perceptual modeling. Bits are then dynamically allocated from a common pool to each coefficient or group of coefficients such that the resulting distortion will be below the masking threshold. This delicate operation will be invalidated if the distortion in one coefficient is affected considerably by the distortions present in other coefficients due to poor sidelobe behavior. Since quantization-noise shaping is the foundation on which audio compression schemes are built, this last reason is actually the most important of the three mentioned. Lower transform coding gain can in many cases be tolerated. Adequate frequency resolution can be provided to the perceptual modeling stage by performing a separate spectral analysis in parallel (as in the MPEG audio coding scheme). A valid foundation for quantization-noise shaping, however, is essential to make the audio compression scheme work.

5.4.2 Synthesis-Filter-Bank Design in Audio Coders

We know from [86] that the magnitude frequency response of the synthesis filters can be approximately given by the magnitude frequency response of the synthesis window modulated to the appropriate frequency. It follows that shaping the frequency response for the synthesis filters is approximately equivalent to shaping the frequency response of the synthesis window. The same can be said of the analysis filters and the analysis window.

One method of achieving the side-lobe behavior necessary for accurate quantization-noise shaping is to increase the number of degrees of freedom by lengthening the synthesis window. If M , the number of subband channels, is held constant, this approach increases K , the overlapping factor. This unfortunately is in conflict with the last design criterion of section 2.2.1, high temporal resolution. If K is a large value, artifacts such as “pre-echos” are exacerbated. It is therefore to our advantage to keep the value of K relatively small. In light of this, we propose that biorthogonality be used to provide enough degrees of freedom to achieve adequate side-lobe attenuation.

Consider the design of the synthesis window for the case of $K = 1$. In the orthogonal case,

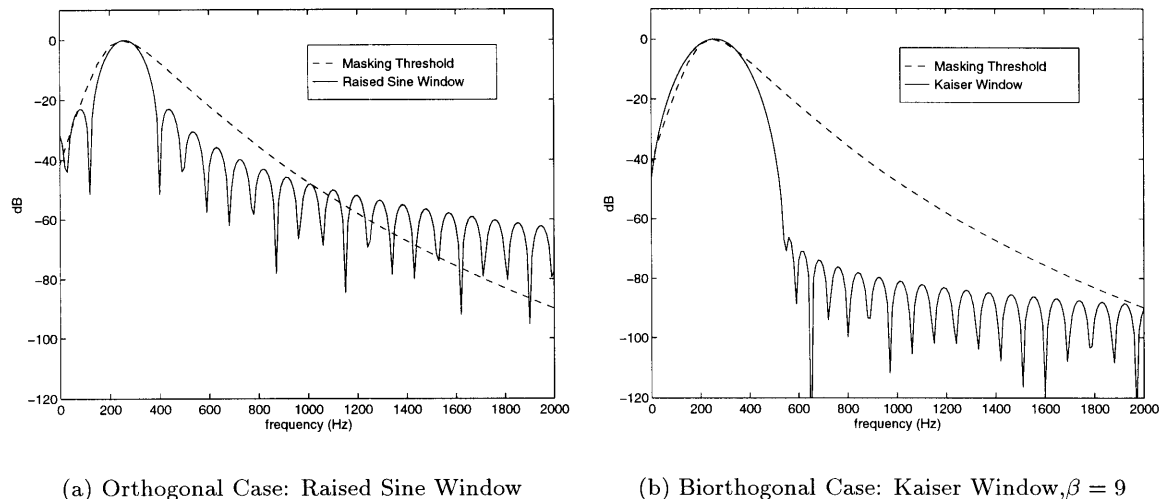


Figure 5.6: Comparison of synthesis windows used for the orthogonal ELT and the biorthogonal cosine-modulate filter bank in relation to the masking threshold of a signal at 250 Hz.

we are essentially limited to the popular choice that satisfies the perfect-reconstruction requirement: the raised-sine window. In figure 5.6(a), we have shown the magnitude frequency response of the window modulated to 250 Hz in relation to a masking threshold derived from a signal at the same frequency. For this, we have assumed a standard sampling rate of 48 kHz so the 512-point window has a duration of 10.7 msec. The masking threshold is computed based on [4]. Note that in the bands close to the signal frequency the sidelobes of the window are sufficiently high to alter substantially the spectral shape of the quantization noise. Unfortunately, the nonlinear nature of the equations in (4.8) makes it difficult for designers to improve upon the analysis window by performing the usual mainlobe-width-versus-sidelobe-attenuation tradeoff. The conventional solution to this problem is to use a conservative perceptual model at the expense of increasing the bit rate.

This situation is rectified in the biorthogonal case. The increase in the number of degrees of freedom to M means that the designer can choose any appropriate symmetric window for the synthesis filter. We have already mentioned the possibility of using the Kaiser window in which the tradeoff between mainlobe width and sidelobe attenuation is controlled by a single parameter β . The relation between the magnitude frequency response of the modulated Kaiser window with $\beta = 9$ and the corresponding masking threshold is shown in figure 5.6(b). The sidelobes in this case are substantially below the masking threshold. Therefore, the effect of leakage on quantization-noise shaping is much reduced and a more aggressive perceptual model can be used. This can potentially improve the performance of the audio coder.

One should bear in mind that the use of a substantially improved synthesis window may lead to poor frequency response for the analysis filters. While this may not affect the perceptual-modeling stage substantially, transform coding gain may be reduced. Under the biorthogonal formulation, we have the flexibility to trade off the two criteria. The optimal amount of tradeoff is an important topic for future study.

Chapter 6

Conclusion

In this dissertation, we have described the current status of audio compression research, placing special emphasis on one important aspect, the short-time spectral decomposition. In particular, we formulated a biorthogonal cosine-modulated filter bank which is a generalization of the ELT. We also showed for three special cases and for the general case that the incorporation of biorthogonality into an M -channel ELT leads to an increase of $M/2$ degrees of freedom. The additional flexibility allows for the design of synthesis filter banks with improved sidelobe behavior. Since the use of cosine-modulated filter banks is common in audio coders, this formulation is of significance to audio researchers.

In order to complete this dissertation, here are some final thoughts on the subjects under investigation.

6.1 On Biorthogonal Transforms & Filter Banks

The formulation of a biorthogonal cosine-modulated filter bank in this thesis directs us to some topics for future study. The most important among these is the development of a better filter-bank design method that can allow control over the design of both the analysis and the synthesis filters. One possible approach is given in section 5.3.2. For this approach to succeed, a better understanding of the optimization techniques is required.

The use of biorthogonal filter banks is clearly not limited to audio compression. In image compression, biorthogonal wavelets have long been used to create linear-phase filters [92]. While the cosine-modulated structure of our biorthogonal formulation precludes linear-phase filters, additional degrees of freedom may be used to advantage. For example, the design flexibility can be used to ensure that the analysis filters for higher subbands have zero DC-gain (equivalent to polyphase normalization). Alternatively, zeros can be placed at the aliasing frequencies of the synthesis filter bank, thereby reducing periodic artifacts at the upsampling grid.

6.2 On Digital Audio Compression

Audio compression by coding in the frequency domain has been the predominant technique in the field for a long time. The latest developments have led to time-varying filter banks and this thesis has suggested the use of biorthogonality. While interesting and constructive, neither time-variation nor biorthogonality is likely to provide a large breakthrough in terms of improved quality or reduced bit rate.

On the not-so-distant horizon, we can perceive the advent of real-time high-quality audio transmission over the internet, or a silicon-memory-based audio player and recorder. Even closer to realization is high-quality audio compression at low bit rates for the MPEG 4 teleconferencing standard. For any of these projects to succeed, a large breakthrough is necessary. Breakthroughs in algorithmic development do not occur through incremental adjustments to existing structures but rather through paradigm shifts. As mentioned earlier in this thesis, conventional wisdom has deemed source modeling impossible for audio signals. The idea, however, warrants further study. At least two major forms of audio signals, speech and music, would benefit from source modeling. The modeling of speech has been a topic of study for many years and researchers have achieved significant gains in compression. If the same can be done to music, source modeling may yet be the edge needed to bring audio compression research into another era.

Bibliography

- [1] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, October 1993.
- [2] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, 1990.
- [3] H. Fletcher, "Auditory patterns," *Review of Modern Physics*, vol. 12, pp. 47 – 65, January 1940.
- [4] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *Journal of the Acoustical Society of America*, vol. 66, pp. 1647–1652, December 1979.
- [5] D. Teh, A. Tan, and S. Koh, "Subband coding of high fidelity quality audio signals at 128 kbps," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II–197–II–200, 1992.
- [6] Y. Sung and J. Yang, "An audio compression system using modified transform coding and dynamic bit allocation," *IEEE Transactions on Consumer Electronics*, vol. 39, pp. 255–259, August 1993.
- [7] L. Fielder and G. Davidson, "AC-2: A family of low complexity transform based music coders," in *AES Workshop on Digital Audio*, (London), October 1991.
- [8] W. Chan and A. Gersho, "High fidelity audio transform coding with vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1109–1112, 1990.
- [9] W. Chan and A. Gersho, "Constrained-storage vector quantization in high fidelity audio transform coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3597–3600, 1991.
- [10] N. Iwakami, T. Moriya, and S. Miki, "High-quality audio-coding at less than 64 KBits/s by using transform-domain weighted interleave vector quantization (TWINVQ)," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3095–3098, May 1995.
- [11] P. A. Monta and S. Cheung, "Low rate audio coder with hierarchical filterbanks and lattice vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, pp. 209–212, April 1994.

- [12] J. D. Johnston and K. Brandenburg, "Wideband coding—perceptual considerations for speech and music," in *Advances in Speech Signal Processing* (S. Furui and M. M. Sondhi, eds.), ch. 4, Marcel Dekker, Inc., 1992.
- [13] X. Lin, R. A. Salami, and R. Steele, "High quality audio coding using analysis-by-synthesis technique," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3617–3620, 1991.
- [14] X. Lin and R. Steele, "Subband coding with modified multipulse LPC for high quality audio," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–201–I–204, 1993.
- [15] S. Boland and M. Deriche, "High quality audio coding using multipulse LPC and wavelet decomposition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 3067–3070, May 1995.
- [16] A. Charbonnier and J. P. Petit, "Sub-band ADPCM coding for high quality," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2540–2543, 1988.
- [17] P. Vörös, "High quality sound coding at 2×64 kbits/s using instantaneous dynamic bit allocation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2536–2539, 1988.
- [18] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using a dynamic dictionary and optimized wavelets," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–197–I–200, 1993.
- [19] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3463–3479, December 1993.
- [20] J. D. Johnston, "Perceptual transform coding of wideband stereo signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1993–1996, 1989.
- [21] W. R. T. ten Kate *et al.*, "Matrixing of bit rate reduced audio signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II–201–II–203, 1992.
- [22] W. R. T. ten Kate, L. M. Van De Kerkhof, and F. F. M. Zijderveld, "A new surround-stereo-sound coding technique," *Journal of the Audio Engineering Society*, vol. 40, pp. 376–382, May 1992.
- [23] K. Brandenburg *et al.*, "Fast signal processor encodes 48 khz/16 bit audio into 3 bit in real time," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2528–2531, 1988.
- [24] K. Brandenburg *et al.*, "Real-time implementation of low complexity transform coding," in *the 84th Convention of the Audio Engineering Society*, (Paris), 1988.

- [25] K. Brandenburg, "OCF — a new coding algorithm for high quality sound signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 141–144, 1987.
- [26] K. Brandenburg *et al.*, "High quality sound coding at 2.5 bit/sample," in *the 84th Convention of the Audio Engineering Society*, (Paris), 1988.
- [27] K. Brandenburg and D. Seitzer, "OCF: Coding high quality audio with data rates of 64 kbit/sec," in *the 85th Convention of the Audio Engineering Society*, (Los Angeles), 1988.
- [28] K. Brandenburg *et al.*, "Low bit rate codecs for audio signals implementation in real time," in *the 85th Convention of the Audio Engineering Society*, (Los Angeles), 1988.
- [29] E. Eberlein, H. Gerhäuser, and S. Krägeloh, "Audio codec for 64 kbit/sec (ISDN channel) — requirements and results," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1105–1108, 1990.
- [30] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-25, pp. 299–309, August 1977.
- [31] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 314–323, February 1988.
- [32] J. D. Johnston, "Estimation of perceptual noise entropy using noise masking criteria," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2524–2527, 1988.
- [33] N. S. Jayant and E. Y. Chen, "Audio compression: Technology and applications," *AT&T Technical Journal*, pp. 23–34, March/April 1995.
- [34] Y. Mahieux, J. P. Petit, and A. Charbonnier, "Transform coding of audio signals using correlation between successive transform blocks," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2021–2024, 1989.
- [35] Y. Mahieux and J. P. Petit, "Transform coding of audio signals at 64 kbits/s," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 0518–0522, 1990.
- [36] Y. Mahieux and J. P. Petit, "High-quality audio transform coding at 64 kps," *IEEE Transactions on Communications*, vol. 42, pp. 3010–3019, November 1994.
- [37] K. Brandenburg, J. Herre, J. D. Johnston, Y. Mahieux, and E. F. Schroeder, "ASPEC: Adaptive spectral entropy coding of high quality music signals," in *the 90th Convention of the Audio Engineering Society*, (Paris), February 1991.
- [38] *MUSICAM—Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing, A Universal Subband Coding System Description*, 1990. ISO/IEC JTC1SC2/WG11, MPEG 90/201.

- [39] International Standards Organization, Moving Picture Experts Group, *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s*. Committee Draft 11172-3, part 3, Audio.
- [40] P. Noll, “Wideband speech and audio coding,” *IEEE Communications Magazine*, vol. 31, pp. 34–44, November 1993.
- [41] R. N. J. Veldhuis, M. Breeuwer, and R. G. van der Waal, “Subband coding of digital audio signals without loss of quality,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2009–2012, 1989.
- [42] R. G. van der Waal and R. N. J. Veldhuis, “Subband coding of stereophonic digital audio signals,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3601–3604, 1991.
- [43] A. Hoogendoorn, “Digital compact cassette,” *Proceedings of the IEEE*, vol. 82, pp. 1479–1489, October 1994.
- [44] T. Yoshida, “The rewritable MiniDisc system,” *Proceedings of the IEEE*, vol. 82, pp. 1492–1500, October 1994.
- [45] R. Dressler, “Dolby pro logic surround sound decoder — principles of operation.” Dolby Laboratories Information, 1988. s89/8540/8624.
- [46] G. Davidson, L. Fielder, and M. Antill, “High quality audio transform coding at 128 kbits/s,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1117–1120, 1990.
- [47] G. Davidson and M. Bosi, “AC-2: High quality digital audio coding for broadcast and storage,” in *NAB 1992 Broadcast Engineering Conference Proceedings*, March 1992.
- [48] M. Davis, “The AC-3 multichannel coder,” in *the 95th Convention of the Audio Engineering Society*, (New York), October 1993.
- [49] “Channel Compatible DigiCipher HDTV System.” Submitted by Massachusetts Institute of Technology, on behalf of the American Television Alliance, May 1992.
- [50] T. Sporer, U. Gbur, J. Herre, and R. Kapust, “Evaluating a measurement system,” *Journal of the Audio Engineering Society*, vol. 43, pp. 353–363, May 1995.
- [51] T. Sporer and K. Brandenburg, “Constraints of filter banks used for perceptual measurement,” *Journal of the Audio Engineering Society*, vol. 43, pp. 107–116, March 1995.
- [52] J. G. Beerends and J. A. Stemerdink, “A perceptual audio quality measure based on a psychoacoustic sound representation,” *Journal of the Audio Engineering Society*, vol. 40, pp. 963–978, December 1992.
- [53] P. P. Vaidyanathan, “Multirate digital filters, filter banks, polyphase networks and applications: A tutorial,” *Proceedings of the IEEE*, vol. 78, pp. 56–93, January 1990.

- [54] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, pp. 14–38, October 1991.
- [55] M. Vetterli and C. Herley, "Wavelet and filter banks: Theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207–2223, September 1992.
- [56] P. P. Vaidyanathan, "Quadrature mirror filter banks M-band extensions and perfect-reconstruction techniques," *IEEE ASSP Magazine*, pp. 4–20, July 1987.
- [57] P. P. Vaidyanathan and P.-Q. Hoang, "The perfect-reconstruction qmf bank: New architectures, solutions and optimization strategies," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2169–2172, 1987.
- [58] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Prentice Hall, 1995.
- [59] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without blocking effects," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 553–559, April 1989.
- [60] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice Hall, 1993.
- [61] Y. F. Dehery, M. Lever, and P. Urcun, "A MUSICAM source codec for digital audio broadcasting and storage," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3605–3608, 1991.
- [62] C. Heegard and T. Shamoan, "High-fidelity audio compression: Fractional-band wavelets," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II–205–II–208, 1992.
- [63] H. S. Malvar, "Extended lapped transforms: Properties, applications, and fast algorithms," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2703–2714, November 1992.
- [64] J. P. Stautner, "High quality audio compression for broadcast and computer applications," in *26th Annual SMPTE Advanced Television and Electronic Imaging Conference*, February 1992.
- [65] L. F. C. Vargas and H. S. Malvar, "ELT-based wavelet coding of high-fidelity audio signals," in *Proceeding of IEEE International Symposium on Circuits and Systems*, (Chicago), pp. 124–127, 1993.
- [66] M. Iwadare, A. Sugiyama, F. Hazu, A. Hirano, and T. Nishitani, "A 128 kb/s hi-fi audio CODEC based on adaptive transform coding with adaptive block size MDCT," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 138–133, January 1992.
- [67] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, pp. 160–175, April 1993.
- [68] C. Herley, J. Kovačević, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3341–3359, December 1993.

- [69] J. Kovačević and M. Vetterli, “Time-varying modulated lapped transforms,” in *Proceedings of the IEEE Asilomar Conference in Signals, System and Computation*, vol. 2, pp. 481–485, 1993.
- [70] R. L. de Queiroz and K. R. Rao, “Time-varying lapped transforms and wavelet packets,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 3293–3305, December 1993.
- [71] I. Sodagar, K. Nayebi, and T. P. Barnwell III, “A class of time-varying wavelet transforms,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, pp. 201–204, 1993.
- [72] I. Sodagar, K. Nayebi, and T. P. Barnwell III, “Time-varying filter banks and wavelets,” *IEEE Transactions on Signal Processing*, vol. 42, pp. 2983–2996, November 1994.
- [73] C. Herley and M. Vetterli, “Orthogonal time-varying filter banks and wavelet packets,” *IEEE Transactions on Signal Processing*, vol. 42, pp. 2650–2663, October 1994.
- [74] C. Herley, “Boundary filters and finite-length signals and time-varying filter banks,” *IEEE Transactions on Circuits and Systems II*, vol. 42, pp. 102–114, February 1995.
- [75] J. Princen and J. D. Johnston, “Audio coding with signal adaptive filterbanks,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 3071–3074, May 1995.
- [76] G. Strang, *Linear Algebra and Its Applications*. Academic Press, 2nd ed., 1980.
- [77] P. P. Vaidyanathan, “Causal FIR matrices with anticausal FIR inverses, and application in characterization of biorthonormal filter banks,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, pp. 177–180, April 1994.
- [78] P. P. Vaidyanathan and T. Chan, “Role of anticausal inverses in multirate filterbanks—part I: System-theoretic fundamentals,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 1090–1102, May 1995.
- [79] P. P. Vaidyanathan and T. Chan, “Role of anticausal inverses in multirate filterbanks—part II: The FIR case, factorizations, and biorthogonal lapped transforms,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 1103–1115, May 1995.
- [80] S. Cheung and J. S. Lim, “Incorporation of biorthogonality into lapped transforms for audio compression,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 3079–3082, May 1995.
- [81] R. D. Koilpillai and P. P. Vaidyanathan, “Cosine-modulated FIR filter banks satisfying perfect reconstruction,” *IEEE Transactions on Signal Processing*, vol. 40, pp. 770–783, April 1992.
- [82] P. Duhamel, Y. Mahieux, and J. P. Petit, “A fast algorithm for the implementation of filter banks based on time domain aliasing cancellation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 2209–2212, May 1991.

- [83] D. Šević and M. Popović, “A new efficient implementation of the oddly stacked Princen-Bradley filter bank,” *IEEE Signal Processing Letters*, vol. 1, pp. 166–168, November 1994.
- [84] D. Šević and M. Popović, “Improved implementation of the Princen-Bradley filter bank,” *IEEE Transactions on Signal Processing*, vol. 42, pp. 3260–3261, November 1994.
- [85] J. P. Princen and A. B. Bradley, “Analysis/synthesis filter bank design based on time domain aliasing cancellation,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 1153–1161, October 1986.
- [86] H. S. Malvar, *Signal Processing with Lapped Transforms*. Artech House, 1991.
- [87] P. P. Vaidyanathan, “Characterization and factorization results for FIR biorthonormal filter banks,” in *Proceedings of the IEEE Asilomar Conference in Signals, System and Computation*, vol. II, pp. 1276–1280, November 1993.
- [88] P. P. Vaidyanathan, “Orthonormal and biorthonormal filter banks as convolvers, and convolutional coding gain,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 2110–2129, June 1993.
- [89] S.-M. Phoong, C. W. Kim, P. P. Vaidyanathan, and R. Ansari, “A new class of two-channel biorthogonal filter banks and wavelet bases,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 649–665, March 1995.
- [90] G. Smart and A. B. Bradley, “Filter bank design based on time domain aliasing cancellation with non-identical windows,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. III, pp. 181–184, April 1994.
- [91] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [92] P. N. Heller, T. Q. Nguyen, H. Singh, and W. K. Carey, “Linear-phase M-band wavelets with application to image coding,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1496–1500, May 1995.